Lipoprotein biogenesis and evolution of lipid transport: an insect perspective

Biogenese van lipoproteïnen en evolutie van lipid transport:
het insecten-systeem als invalshoek

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht
op gezag van de rector magnificus, prof. dr. W.H. Gispen,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op
dinsdag 25 april 2006 des middags te 12.45 uur
door

Marcel Martinus Maria Wilhelmus Smolenaars

geboren op 20 april 1978, te Weert

Promotor: Prof. Dr. D.J. van der Horst

Co-promotor: Dr. K.W. Rodenburg

Lipoprotein biogenesis and evolution of lipid transport: an insect perspective

Thesis committee:

Dr. J. Bogerd
(Universiteit Utrecht)

Prof. Dr. M.R. Egmond
(Universiteit Utrecht)

Prof. Dr. J.F.C. Glatz
(Universiteit Maastricht)

Prof. Dr. K.W.A. Wirtz
(Universiteit Utrecht)

Paranimfen:

Antoine de Morrée
Roland Geraerts

The research described in this thesis was performed at the Biochemical Physiology research group within the Department of Biology and Institute of Biomembranes, Utrecht University, The Netherlands.

ISBN-10: 90-393-4226-1
ISBN-13: 978-90-393-4226-8
Printing: Ponsen & Looijen B.V., Wageningen, The Netherlands
Book design: Marcel Smolenaars and Klaas van Gisbergen

Contents

Abbreviations

aa	amino acids
apoB	apolipoprotein B
apoER2	apolipoprotein E receptor 2
apoLp	apolipophorin
ARP	apolipophorin-II/I-related protein
cA/GDPK	cAMP/cGMP dependent protein kinase
CE	cholesteryl esters
CK2	casein kinase-2
CP	clotting protein
DAG	diacylglycerol
decRVKRcmk	decanoyl-Arg-Val-Lys-Arg-chloromethyl ketone
dMTP	*Drosophila melanogaster* microsomal triglyceride transfer protein
EGF	epidermal growth factor
ER	endoplasmic reticulum
HDL	high-density lipoprotein
HDLp	high-density lipophorin
ICD	intracellular domain
LDL	low-density lipoprotein
LDLp	low-density lipophorin
LDLR	low-density lipoprotein receptor
LLT	large lipid transfer
LLTP	large lipid transfer proteins
Lp	lipophorin
LpR	lipophorin receptor
LRP	LDLR related protein
LTP	lipid transfer particle
MEP	melanin-engaging protein
ML	maximum likelihood
MTP	microsomal triglyceride transfer protein

OLGD	O-linked glycosylation domain
PC	proprotein convertase
PDI	protein disulfide isomerase
PKC	protein kinase C
SOD	superoxide dismutase
TAG	triacylglycerol
TMD	transmembrane domain
VGR	vitellogenin receptor
VLDL	very low-density lipoprotein
Vtg	vitellogenin
vWF-D	von Willebrand Factor D

Chapter 1

Introduction

Lipids are essential to organisms as they constitute the major building materials of membranes, function as second messengers in signalling events, serve as an energy source and are stored as an energy reserve. As lipids are poorly soluble in water due to their hydrophobic nature, multicellular organisms require a dedicated transport system. Animals developed a unique solution to accomplish lipid transport in their circulation, namely lipoproteins. These macromolecular particles, consisting of one or more proteins and up to hundreds of lipid molecules, have been identified throughout the animal kingdom (Figure 1), from invertebrate phyla such as Nematoda (Blumenthal *et al.*, 1984), Mollusca (Heras and Pollera, 2002), Cnidaria (Tidball, 1981), Annelida (Schenk *et al.*, 2006), Arthropoda (for review see Van der Horst *et al.*, 2002) and Echinodermata (Marsh, 1968), to vertebrates (for review see Chapman *et al.*, 1980).

Generally, lipoproteins are globular particles that consist of an inner core of hydrophobic lipids (*e.g.* triacylglycerol (TAG) and cholesteryl esters (CE)) and an outer shell of amphipathic lipids (*e.g.* phospholipids) interspersed with amphipathic protein components (for reviews see Hevanoja *et al.*, 2000; Segrest *et al.*, 2001). In addition, disc-shaped (precursor) lipoproteins have been observed that are constituted of a bilayer surrounded by amphipathic protein components that cover the fatty acid chains exposed at the circumference (for review see Davidson and Silva, 2005). Both globular and disc-shaped arrangements solubilize lipids, as the amphipathic outer structures cover the hydrophobic-hydrophilic interface, arranged with their hydrophobic sides facing the hydrophobic core and their hydrophylic sides facing the aqueous medium. The lipoprotein proteins, apolipoproteins, associate to hydrophobic lipid surfaces by means of amphipathic α-helix and amphipathic β-sheet structures. Non-exchangeable apolipoproteins remain permanently associated with the lipoprotein, whereas exchangeable apolipoproteins are able to dissociate and re-associate. Lipoproteins are usually classified according to their buoyant density range, expressed in gram per ml, as a relatively higher lipid content results in a decrease of the density of a lipoprotein.

Lipoproteins and their role in lipid transport have been studied extensively in mammals, because of their involvement in human disorders such as atherosclerosis, diabetes mellitus type 2, and obesity (for reviews see Olofsson and Borén, 2005; Shoulders and Shelness, 2005; Taskinen, 2005). Understanding structure and functioning of alternate lipoprotein systems in other animal species may offer novel perspectives on human lipoprotein function and may eventually provide novel treatments of human lipoprotein-related disorders. In this respect, particularly insect lipoproteins –the structure

and function of which have been relatively well studied compared to other invertebrate classes– may prove relevant, as previous investigations have revealed important similarities as well as interesting differences between insect and mammalian lipoprotein systems (for reviews see Van der Horst *et al.*, 2002; Rodenburg and Van der Horst, 2005; Martins and Redgrave, 2004). These differences may result from modification of the

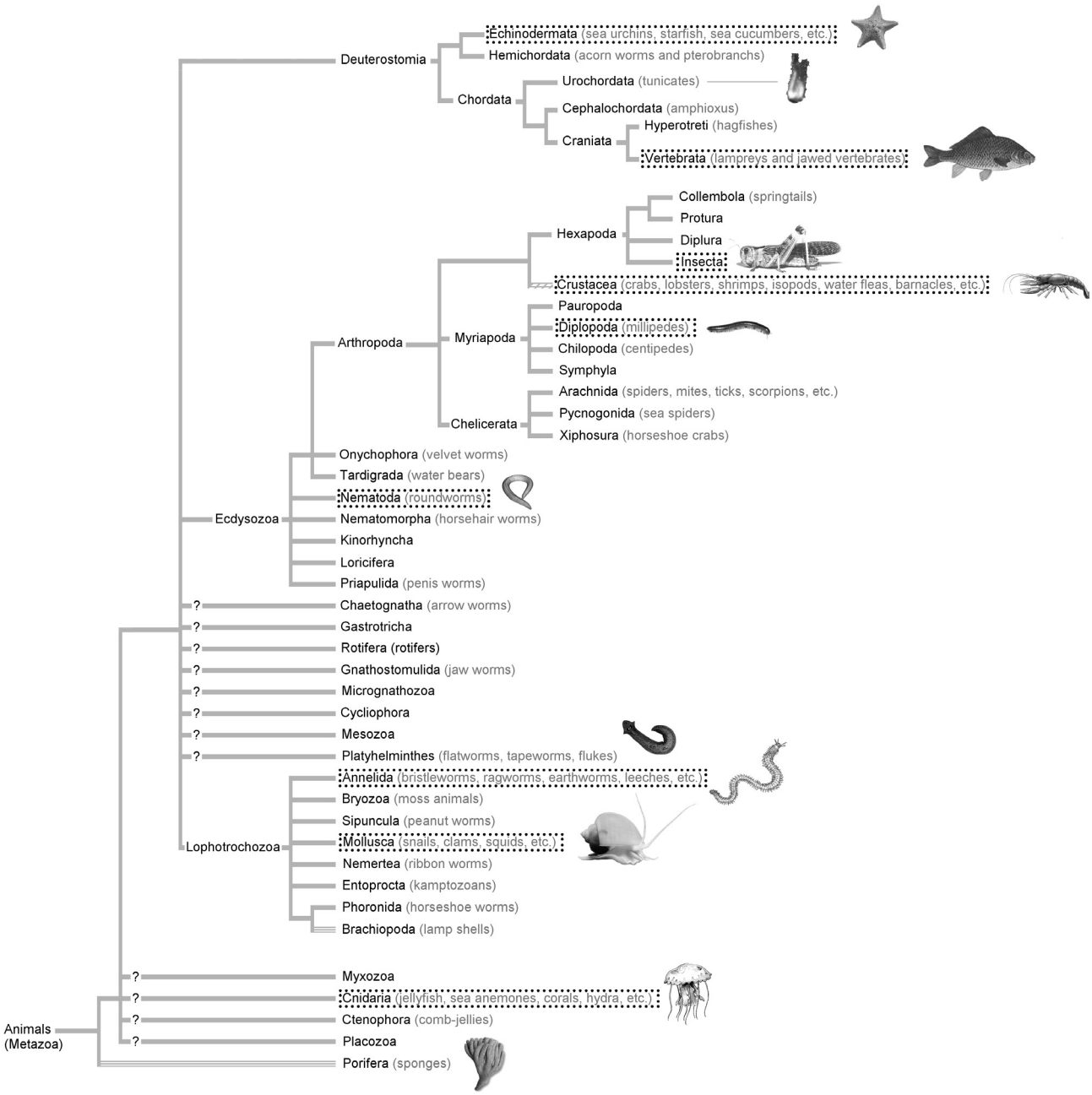


Figure 1. Lipoproteins across the animal kingdom. A phylogenetic tree illustrating the relations among recognized extant animal phyla (Maddison and Schulz, 2004). Phyla in which lipoproteins have been described are indicated by dotted boxes. Question marks indicate groups of uncertain phylogenetic position (*e.g.* Platyhelminthes), whereas groups that are or may be polyphyletic are indicated by branches constituted by three parallel lines (*e.g.* Porifera) or a hatched line (Crustacea), respectively. The groups of Deuterostomia and Arthropoda have been expanded to illustrate the position of vertebrates and insects, respectively.

structural components involved, during evolution as well as lipoprotein biosynthesis. From this perspective, this thesis presents work on the biosynthesis of insect lipoprotein, the evolutionary origin and diversification of lipoprotein precursor proteins, and sequence properties of lipoprotein receptors.

Insect lipoproteins

Lipoprotein-mediated lipid transport in insects involves a single multifunctional carrier particle that was named lipophorin, according to the Greek words *lipos*, *i.e.* lipid, and *phoros*, *i.e.* bearing (Chino *et al.*, 1981). Lipophorin is synthesized in the insect fat body and present in large amounts in the hemolymph (Weers *et al.*, 1992). It consists of two protein components, apolipophorin-I and apolipophorin–II (apoLp-I and apoLp-II), with molecular masses of ~250 and ~75 kDa, respectively (Shapiro *et al.*, 1984; Beenakkers *et al.*, 1985; Robbs *et al.*, 1985; Weers *et al.*, 1993). Using density gradient ultracentrifugation, the density of lipophorin in resting insects was determined to be 1.09-1.18 g/ml, depending on the insect species studied (for review see Soulages and Wells, 1994). Thus, according to the density-based lipoprotein classification system of mammals, the resting state lipophorin is a high-density lipoprotein (HDLp). Although insects store lipids in the fat body as TAG, the most prevalent lipid in the HDLp of most insects is diacylglycerol (DAG), in addition to phospholipids, sterols, hydrocarbons, and limited amounts of TAG (for reviews see Van der Horst, 1990; Van der Horst *et al.*, 1993; Soulages and Wells, 1994).

The function of lipophorin in lipid transport has particularly been studied in the context of intense muscular activity (for reviews see Van der Horst, 1990; Ryan and Van der Horst, 2000; Van der Horst *et al.*, 2001; Van der Horst and Ryan, 2004), especially during migratory flight of insects such as the migratory locust, *Locusta migratoria* (Figure 2). Similar to sustained running in humans, the locust switches from carbohydrates to lipids as its major energy source for muscular activity upon sustained flight. However, in contrast to humans, in which lipid (free fatty acids) mobilized from adipose tissue is transported bound to albumin, in insects the lipoprotein system mediates lipid transport through the circulation. Lipids are mobilized from the major insect adipose tissue, the fat body, and transported as DAG by lipophorin via the blood (hemolymph) to the flight muscles (Figure 3). HDLp acquires a large amount of DAG at the plasma membrane of the fat

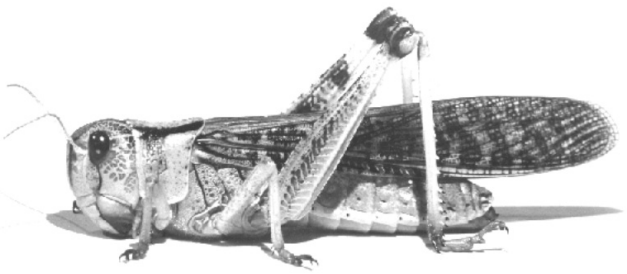


Figure 2. *Locusta migratoria*, the African migratory locust.

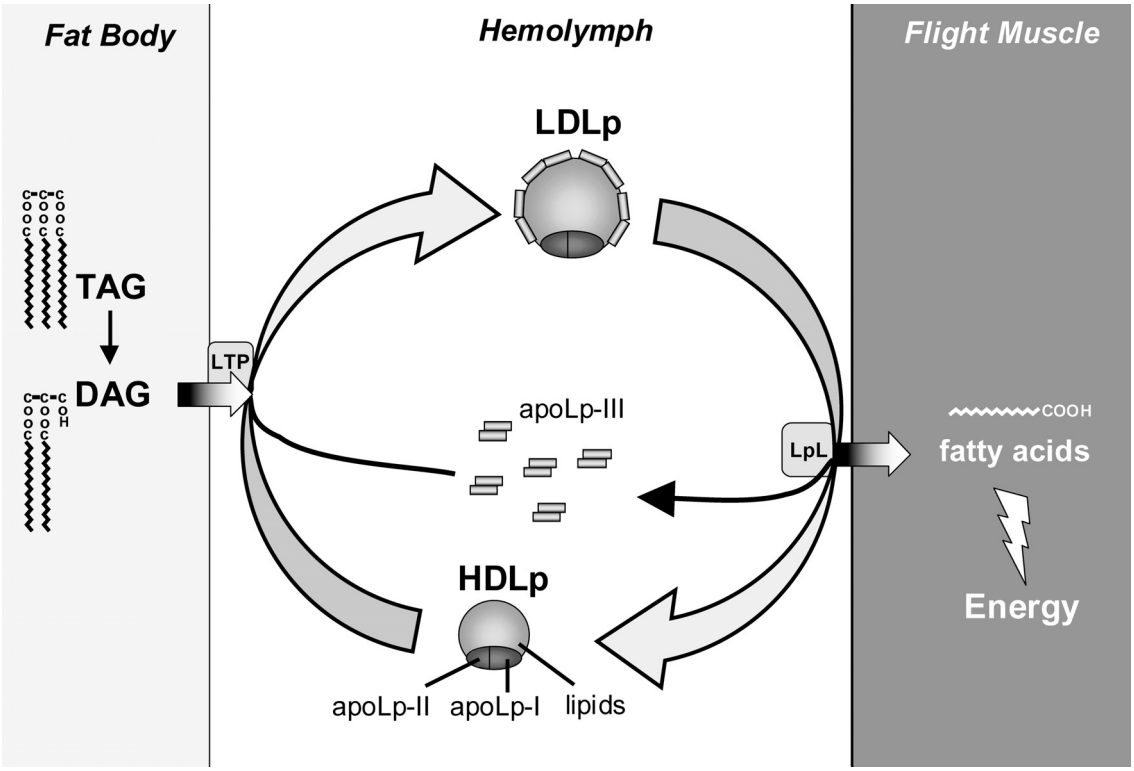


Figure 3. The lipoprotein shuttle system during insect flight. Sustained flight is powered by lipids that are mobilized in the form of DAG from cytosolic TAG stores in the fat body. HDLp attains DAG from fat body cells; concomitantly, several apoLp-III associate with the lipid-enriched particle, resulting in the transformation of HDLp to a LDLp particle. The transfer of DAG onto HDLp at the fat body is mediated by the activity of a lipid transfer particle (LTP). DAG hydrolysis by a flight muscle lipophorin lipase (LpL) results in the release of apoLp-III and finally in recovery of HDLp, that can be lipid loaded at the fat body again. Accordingly, lipids are transported from the fat body to the flight muscle by a mechanism involving a reusable lipoprotein.

body cells. Concomitant with lipid loading, multiple copies (16 to 18) of the exchangeable apolipoprotein apoLp-III reversibly associate to the lipophorin surface by means of their amphipathic α-helices, accordingly stabilizing the lipid load of lipophorin. As a result, HDLp is transformed to a lower density lipoprotein, which in accordance with the mammalian density-based classification system is termed low-density lipophorin (LDLp), with a buoyant density of ~1.06 g/ml. At the flight muscle surface, the lipid cargo of LDLp is decreased as lipophorin lipase activity results in hydrolysis of bound DAG, leading to the dissociation of apoLp-III and the recovery of HDLp, that is once more available for lipid uptake and LDLp formation. Thus, lipophorin functions as a reusable shuttle for lipid transport during sustained insect flight, whereas apoLp-III acts as a lipid-triggered molecular switch in this efficient mechanism (for review see Weers and Ryan, 2003).

In contrast to the formation of LDLp during muscular activity, dietary lipid uptake and resting-state lipid homeostasis do not involve major lipophorin transformation, and may be mediated by minor loading and extraction of lipids to HDLp (for reviews see Van

der Horst and Ryan, 2004; Rodenburg and Van der Horst, 2005), sometimes possibly associated with the binding or dissociation of one or a few copies of apoLp-III. In addition to its functioning during rest and flight activity, lipophorin has been implicated in vitellogenesis, *i.e.* the supply of large amounts nutrients to the developing oocyte in oviparous animals, such as insects, as oocytes endocytose and accumulate HDLp as a yolk protein, and may additionally acquire lipids from LDLp (Kawooya and Law, 1988; Kawooya *et al.*, 1988; Sun *et al.*, 2000).

The mechanism of lipid transfer between lipophorin and tissues is not completely understood, but involves a lipid transfer particle (LTP), a very high-density lipoprotein (1.23 g/ml) of ~670 kDa that consists of three different apoproteins in addition to ~14% lipids (for reviews see Ryan and Van der Horst, 2000; Van der Horst and Ryan, 2004). In addition, lipophorin receptors have been characterized at the plasma membrane of lipophorin-binding tissues such as flight muscle and fat body, that may assist in lipid transfer between lipoproteins and cells (or *vice versa*) (Van Antwerpen *et al.*, 1988; Dantuma *et al.*, 1996, 1999; Van Hoof *et al.*, 2002, 2003).

Mammalian lipoproteins

Unlike insects, mammals rely on an array of lipoproteins from different sources. Three major interconnected pathways of lipoprotein-mediated lipid transport have been recognized (Montgomery *et al.*, 1996; Frayn, 2003) (detailed in Figure 4). First, the enterocytes of the small intestine secrete chylomicrons to distribute dietary lipids to the rest of the body (exogenous lipid transport). Second, the liver distributes (esterified) cholesterol and TAG to extrahepatic tissues via the production of very low-density lipoprotein (VLDL) (endogenous lipid transport). Third, excess cholesterol from peripheral tissues is redistributed by high-density lipoprotein (HDL) (reverse cholesterol transport; for review see Assmann and Nofer, 2003). Whereas endogenous and exogenous lipid transport depend on *de novo* biosynthesis and subsequent degradation of lipoproteins, HDL is continuously regenerated to undertake its function of inducing cholesterol efflux and transporting cholesterol (for review see Assmann and Nofer, 2003).

HDL is composed of exchangeable apolipoproteins only, particularly apolipoprotein A-I, and a relatively low amount of lipids, mostly phospolipids, cholesterol, and cholesteryl esters. In constrast, the lipoproteins of the endogenous and exogenous lipid transport pathways (VLDL, chylomicrons and their derived lipoprotein particles) are stabilized by a single non-exchangeable apolipoprotein, apolipoprotein B (apoB). Mammalian species have been found to express apoB in two variants, *i.e.* the complete encoded protein (apoB-100, in humans present in VLDL) as well as a C-terminal apoB truncation variant corresponding to the N-terminal 48% of the total apoB (apoB-48, in humans present in chylomicrons). The apoB-48 variant is generated by modification of a single nucleotide in the apoB mRNA by a dedicated RNA editing factor,

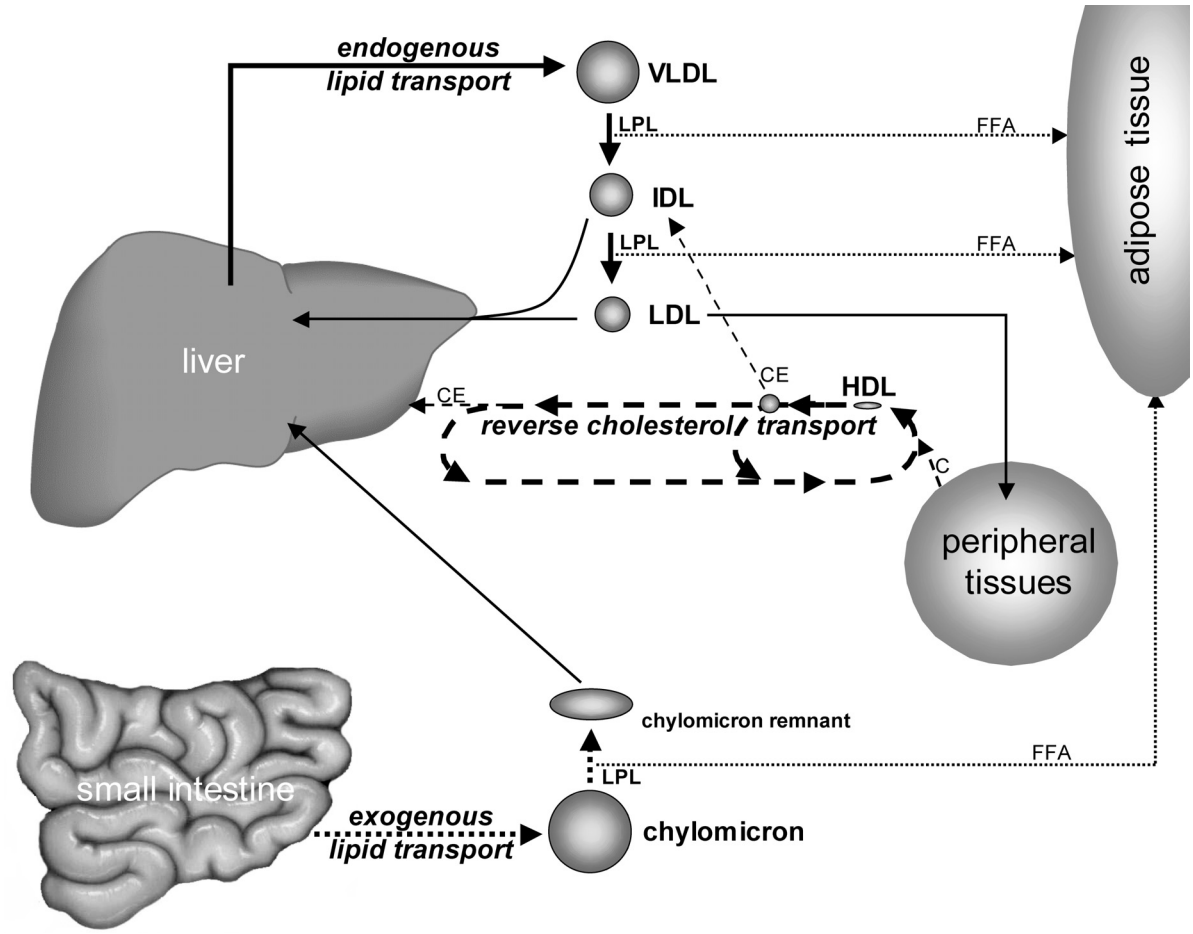


Figure 4. Schematic overview of lipoprotein-mediated lipid transport in mammals. Three major lipoprotein pathways are recognized. Exogenous lipid transport refers to the uptake of dietary fats via the biosynthesis of chylomicron lipoproteins by the small intestine (dotted thick lines). Endogenous lipid transport mediates the distribution of (partly *de novo* synthesized) lipids by the liver via the biosynthesis of VLDL (solid thick lines). Lipids from chylomicrons and VLDL are hydrolyzed by lipoprotein lipase (LpL). Whereas the resulting free fatty acids (FFA) are taken up by adipose tissues (dotted thin lines) and to a lesser extent by peripheral tissues, the resulting higher density lipoprotein remnants are taken up and degraded by tissues (solid thin lines), *i.e.* chylomicron remnants are taken up by the liver, and LDL by the liver as well as peripheral tissues. The third major lipoprotein pathway, reverse cholesterol transport, involves the transfer of cholesterol (C) from peripheral tissues to the liver via HDL (dashed thick lines). Following uptake from peripheral tissues, cholesterol in HDL is converted to cholesteryl esters (CE) that are transferred (dashed thin lines) either directly to the liver or incorporated into IDL (and accordingly LDL). Whereas endogenous and exogenous lipid transport depend on the biosynthesis of novel lipoprotein particles, HDL is continuously regenerated while undertaking its function of transporting cholesterol from peripheral tissues to the liver.

resulting in the introduction of an early stop codon (for review see Davidson and Shelness, 2000). Multiple exchangeable apolipoproteins may bind to these apoB lipoproteins and accordingly direct their function and fate. Upon biosynthesis, the main lipid component of apoB lipoproteins is TAG.

Table 1: Comparison of lipoproteins in insects and mammals.

Insects	Mammals
Lipophorin	VLDL, chylomicron
- general-purpose lipoprotein	- generally synthesized after dietary lipid intake
- reusable shuttle mechanism, lipid (un)loading at the plasma membrane	- not reusable: lipid transport driven by *de novo* biosynthesis of lipoproteins and post-endocytic degradation of remnant particles
- 2 non-exchangeable apolipoproteins (apoLp-I and apoLp-II)	- 1 non-exchangeable apolipoprotein (apoB-100 or apoB-48)
- DAG as major neutral lipid	- TAG as major neutral lipid of chylomicrons and VLDL
- enhanced lipid binding via apoLp-III association	HDL
	- exchangeable apoA-I as major apolipoprotein
	- PL and CE are major lipids
	- reusable: lipid transport without degradation of HDL

Insect and mammalian lipoprotein systems compared

Despite common principles, the lipoproteins of insects and mammals show marked differences in components and function (Table 1) that may result from the modification of ancient structural components. Remarkably, the two non-exchangeable apolipoproteins in insect lipoprotein, apoLp-I and apoLp-II, were found to derive from cleavage of their common precursor protein, apolipophorin-II/I (apoLp-II/I), that was demonstrated to be a homolog of the single non-exchangeable apolipoprotein of mammalian lipoproteins, apoB (Weers *et al.*, 1993; Babin *et al.*, 1999; Mann *et al.*, 1999). The notion that insects and mammals rely on common structures for lipid transport is further supported by the involvement in both taxa of members of the low-density lipoprotein receptor (LDLR) family in lipoprotein binding (for reviews see Sappington and Raikhel, 1998a; Van der Horst *et al.*, 2002; Rodenburg and Van der Horst, 2005). Thus, differences between insect and mammalian lipoprotein structure and function in part reflect the modification of ancient structural components.

Lipophorin biosynthesis in insects

The common precursor, from which apoLp-I and apoLp-II are derived by post-translational cleavage (Weers *et al.*, 1993), was named apoLp-II/I, as molecular characterization showed that the protein is arranged with apoLp-II at the N-terminal end and apoLp-I at the C-terminal end (Kutty *et al.*, 1996; Sundermeyer *et al.*, 1996; Bogerd *et al.*, 2000). Cleavage of apoLp-II/I can be related to the activity of furin, a member of the proprotein convertase (PC) family of subtilisin-like serine endoproteases (Taylor *et al.*, 2003), as *L. migratoria* apoLp-II/I is cleaved immediately C-terminal to a consensus

substrate sequence for furin, R-X-K/R-R, that is also present in the other characterized insect apoLp-II/I (Kutty *et al.*, 1996; Sundermeyer *et al.*, 1996; Bogerd *et al.*, 2000).

The major site of lipophorin biosynthesis is the fat body (analogous to mammalian adipose tissue and liver combined) that secretes apoLp-I and apoLp-II together in a lipophorin particle with a buoyant density in the high-density or very high-density lipoprotein range, dependent on the species (Prasad *et al.*, 1986; Venkatesh *et al.*, 1987; Capurro and De Bianchi, 1990; Weers *et al.*, 1992; Van Heusden *et al.*, 1998), containing far less and different (DAG rather than TAG) associated lipids than apoB lipoproteins. It remains to be established whether lipidation starts prior to cleavage of apoLp-II/I, and whether lipidation is completed intracellularly. LTP has been suggested to complete lipidation at the fat body plasma membrane (Capurro and De Bianchi, 1990). Thus, insect lipoprotein biosynthesis differs from mammalian apoB lipoprotein formation in the amount and type of lipid acquired, cleavage of the structural apolipoprotein, and possibly also in the localization of lipidation. Nonetheless, homology between apoLp-II/I and apoB (Babin *et al.*, 1999; Mann *et al.*, 1999), suggests similarities in their mechanisms of lipid acquisition.

ApoB lipoprotein biosynthesis in mammals

Compared to the biosynthesis of lipophorin, the biogenesis of apoB to a lipoprotein has been studied extensively at both the molecular and cellular level. This process occurs in the secretory pathway of the cell, and is proposed to consist of co- and post-translational steps (Figure 5; for reviews see Shelness and Sellers, 2001; Mahmood Hussain *et al.*, 2001; Olofsson and Borén, 2005). ApoB lipoprotein biogenesis starts upon translation of apoB mRNA and translocation of the nascent polypeptide into the ER. Following its translocation, the N-terminal ~900 amino acids of apoB fold into the large lipid transfer (LLT) module that constitutes a lipid-binding cavity and interacts with microsomal triglyceride transfer protein (MTP). MTP facilitates and likely mediates lipid transfer to the nascent apoB polypeptide, and newly translocated amphipathic β-strands and α-helices in apoB associate to and stabilize the growing lipid core. MTP is retained in the secretory pathway due to its stable interaction with the ER-resident chaperone protein disulfide isomerase (PDI) that is also essential to the lipid transfer activity and solubility of MTP (Wetterau *et al.*, 1990; Wetterau *et al.*, 1991; Lamberg *et al.*, 1996). Accordingly, misfolding of apoB and its subsequent intracellular retention and degradation are prevented. Following apoB translation and translocation, the lipoprotein particle continues along the secretory pathway and acquires the bulk of its neutral lipids (TAG), possibly by fusion of an intralumenal neutral lipid droplet to the primordial lipoprotein (for reviews see Shelness and Sellers, 2001; Mahmood Hussain *et al.*, 2001; Olofsson and Borén, 2005).

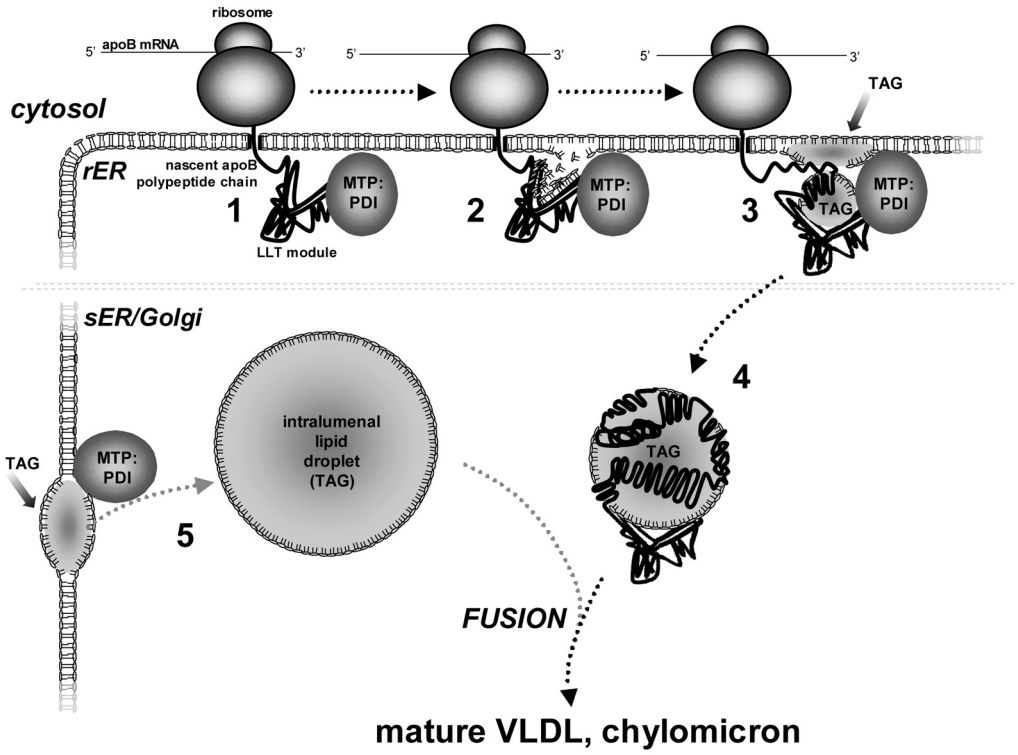


Figure 5: Schematic overview of the biogenesis of chylomicron and VLDL in mammals. The two major steps in the intracellular biosynthesis of these apoB lipoproteins are illustrated by five numbered stages. In the first step, apoB is translated and translocated into the lumen of the rough endoplasmic reticulum (rER), and concomitantly lipidated. Upon its translocation (top left), the N-terminal LLT module of apoB folds and associates with MTP (1), that forms a complex with PDI, resulting in the co-translational lipidation of apoB (2, 3). Whereas the LLT module can accumulate phospholipids (2), acquisition of TAG requires the C-terminal amphipathic clusters to stabilize the nascent TAG lipid core (3). Following translation and translocation, the primordial apoB lipoprotein continues along the secretory pathway (4). In the smooth ER (sER) or possibly the Golgi apparatus, it acquires most of its neutral lipids in a second step that has been proposed to involve the fusion with a pre-existing intralumenal lipid droplet (5). MTP may participate in the accretion of such lipid droplets. The transfer of individual phospholipids out of the ER membrane and the accumulation of TAG in between the membrane bilayer are hypothetical.

ApoB can associate hundreds of lipid molecules by means of amphipathic structures that envelope and stabilize a core of lipids, hence shielding hydrophobic lipids from the hydrophilic environment (Hevonoja *et al.*, 2000; Segrest *et al.*, 2001). Five clusters enriched in either amphipathic α-helices or amphipathic β-strands are predicted to be present in apoB, organized along the polypeptide as N-α_1-β_1-α_2-β_2-α_3-C (Segrest *et al.*, 1994, 1998; Figure 6). The α_1 cluster and the N-terminal part of the β_1 cluster constitute the LLT module. The C-terminal β_1-α_2-β_2-α_3 clusters stabilize expansion of the initial lipid core in the LLT module and actually possess most lipid-binding capacity (Segrest *et al.*, 2001). The β_1 cluster appears to be especially important to buoyant lipoprotein formation, as illustrated by the ability of apoB-48, containing little more than the α_1 and

β_1 clusters (Figure 6), to constitute the extremely TAG-rich chylomicrons. Thus, mammalian lipoprotein assembly is enabled by the molecular architecture of apoB as well as the role of MTP.

Metazoan proteins involved in lipid transport: the LLTP family

Mammalian apoB and insect apoLp-II/I are homologous members of the family of large lipid transfer proteins (LLTP) that also includes vitellogenin (Vtg) and microsomal triglyceride transfer protein (MTP) (Figure 6; Babin *et al.*, 1999). All LLTP bind lipids, albeit in different amounts and for different purposes. Vtg is the major yolk protein found in females of most egg-laying animals, vertebrates as well as invertebrates, supplying the developing oocyte with nutrients, including lipids. Vtg forms a very high-density lipoprotein and accordingly binds a limited amount of lipids, mostly phospholipids. MTP is present in vertebrates and invertebrates and facilitates the biosynthesis of its LLTP family members apoB as well as Vtg (Shibata *et al.*, 2003; Sellers *et al.*, 2005; for reviews see Shelness and Ledford, 2005; Shoulders and Shelness, 2005).

LLTP are defined by the presence of a large lipid transfer (LLT) module, comprising the N-terminal ~900 amino acids (aa) with discernable sequence and structure conservation among LLTP (Raag *et al.*, 1988; Anderson *et al.*, 1998; Babin *et al.*, 1999; Mann *et al.*, 1999; Segrest *et al.*, 1999; Read *et al.*, 2000; Thompson and Banaszak, 2002; Richardson *et al.*, 2005). The N-terminal ~600 residues form a barrel-like β-sheet domain (βC) and a coiled, horsehoe-shaped α-helical domain that corresponds to the α_1 amphipathic cluster. The C-terminal portions of the LLT module fold into two amphipathic β sheets (βA and βB) that constitute the opposite sides of a lipid-binding cavity. Studies on apoB lipoprotein biosynthesis indicate that the LLT module in apoB provides the structural basis for the initial acquisition and binding of lipids (for reviews see Shelness and Sellers, 2001; Mahmood Hussain *et al.*, 2001; Olofsson and Borén, 2005). Nonetheless, the exact role of the interaction between the LLT modules of apoB and MTP in apoB lipoprotein biogenesis remains elusive.

MTP is the smallest LLTP and contains only little more sequence than the LLT module itself. In contrast, vertebrate apoB spans ~4500 aa, insect apoLp-II/I ~3350 aa, and Vtg ~1700 aa (Figure 6). In addition to the LLT module, Vtg and apoLp-II/I contain a single von Willebrand Factor D (vWF-D) module of ~150 aa near their C-terminus, unlike MTP and apoB (Babin *et al.*, 1999; Hall *et al.*, 1999; Yeh *et al.*, 1999). The function of this domain in these proteins is unknown as yet. Apart from the LLT and vWF-D modules, hardly any sequence similarity can be observed among apoB, apoLp-II/I, Vtg and MTP (Babin *et al.*, 1999), and it remains to be established whether amphipathic clusters as in apoB, organized as N-α_1-β_1-α_2-β_2-α_3-C, are also present in the other LLTP. Importantly, the evolutionary relations between these LLTP remain to be established (Babin *et al.*, 1999).

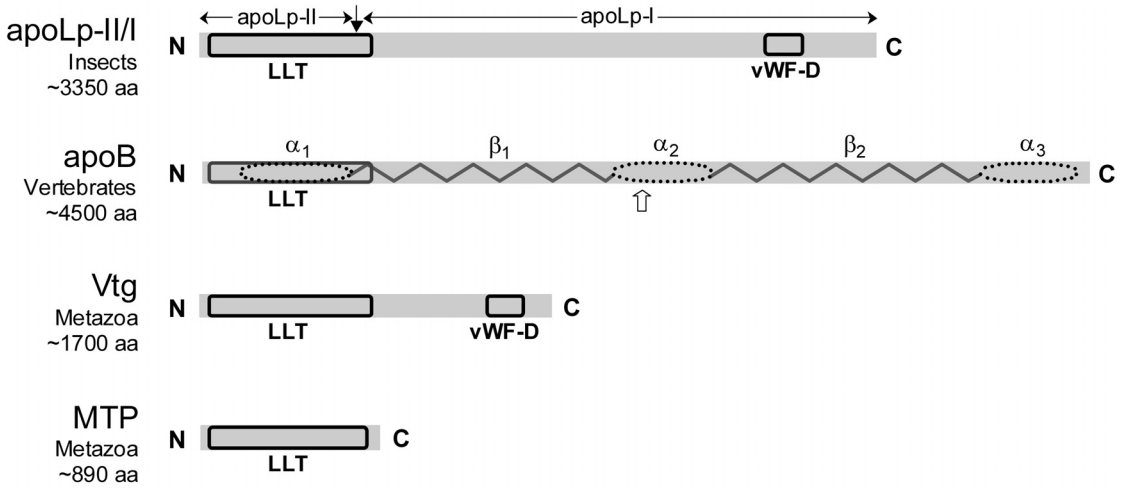


Figure 6. Structural organisation of LLTP family members. The name, phylogenetic distribution and average size in amino acids (aa) is shown for each member. Members are represented by grey boxes, sized proportional to the average protein amino acid length, in which the recognized large lipid transfer (LLT) and von Willebrand Factor D (vWF-D) modules are indicated at their positions by black boxes. The conserved location of the cleavage site in apoLp-II/I is indicated by a closed arrow. For apoB, the organization of amphipathic clusters is indicated by dark grey zig-zag lines representing clusters enriched in amphipathic β-strands, and dashed ovals representing clusters enriched in amphipathic α-strands. Human chylomicrons contain a truncated variant of apoB, apoB-48, as a result of the introduction of an early stop codon in apoB mRNA by RNA editing, and the C-terminal end of apoB-48 is indicated by an open arrow. The N-terminal signal peptide, present in all LLTP, has been omitted.

Metazoan proteins involved in lipid transport: the LDLR family

The LLTP family members apoB, apoLp-II/I and Vtg of distinct animal phyla have been demonstrated to bind to a family of homologous receptors, the low-density lipoprotein receptor (LDLR) family (for reviews see Sappington and Raikhel, 1998a; Herz and Bock, 2002; Schneider and Nimpf, 2003; Jeon and Blacklow, 2005). Accordingly, the LLTP and LDLR families seem to present an example of co-evolution of ligands and receptors, although LDLR family members can bind many additional ligands (see May *et al.*, 2005). LDLR family members were initially characterized as ligand-endocytosing receptors. At present, intracellular signalling functions have also been ascribed to several family members (for review see May *et al.*, 2005).

The archetypical LDLR family member, human LDLR, is composed of five distinct functional domains (Figure 7), from N- to C-terminus: (1) the ligand-binding domain that consist of multiple consecutive imperfect cysteine-rich repeats and binds diverse ligands, including specific LLTP members, (2) the epidermal growth factor (EGF) precursor homology domain, composed of three EGF precursor repeats and a β-propeller structure, that functions in dissociaton of ligand after endocytic uptake, (3) the O-linked glycosylation domain that may be heavily glycosylated and has been suggested to function

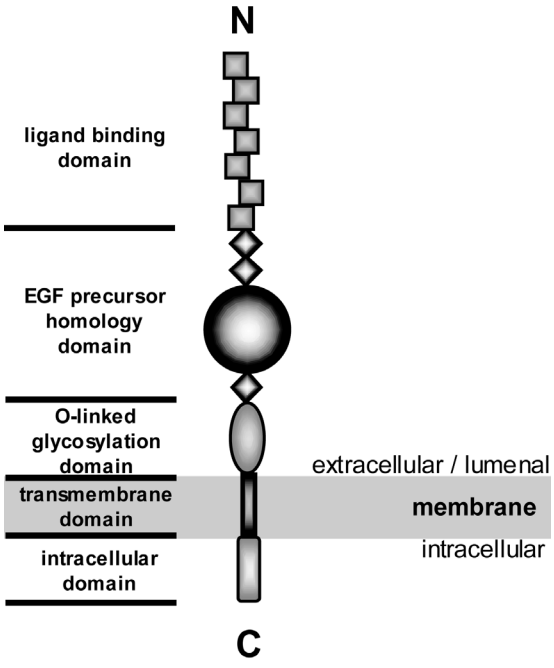


Figure 7. Structural organisation of the human LDLR. The receptor contains a ligand-binding domain composed of 7 cysteine-rich repeats (squares); an EGF homology precursor domain comprising two consecutive EGF repeats (diamonds) that are separated from a third EGF repeat by a β-propeller structure (circle); an O-linked glycosylation domain (oval); a transmembrane domain (narrow rectangle); and an intracellular cytoplasmic tail (broad rectangle).

as a spacer, and to render stability, (4) the transmembrane domain that anchors the receptor in the membrane, and (5) the intracellular tail that plays an essential role in endocytic uptake of the receptor and bound ligands, and is most likely also involved in intracellular signaling. Unlike the N-terminal ligand-binding and EGF precursor homology domains, the three C-terminal domains are composed of non-recurring sequences.

In addition to LDLR, mammals express several other LDLR family members, including VLDL receptor (VLDLR), LDLR-related protein 1 (LRP1), LRP2, and LRP8 (for reviews see Schneider and Nimpf, 2003; May *et al.*, 2005; Jeon and Blacklow, 2005). These receptors contain the same domains as present in LDLR, with the exception of the O-linked glycosylation domain that is absent from LRP1 and LRP2. However, several family members (*e.g.* LRP1 and LRP2) contain multiple ligand-binding and EGF precursor homology domains, or differ in the amount of repeats within these domains (*e.g.* LDLR has 7 cysteine-rich repeats in its single ligand-binding domain, whereas VLDLR contains 8) (for reviews see Herz and Bock, 2002; Schneider and Nimpf, 2003; Jeon and Blacklow, 2005). LDLR family members have also been identified in insects, displaying affinity for lipophorin (Dantuma *et al.*, 1999; Cheon *et al.*, 2001; Lee *et al.*, 2003; Seo *et*

al., 2003) or vitellogenin (Sappington *et al.*, 1996; for review see Sappington and Raikhel, 1998a).

Most LDLR are endocytosed via clathrin-coated pits, due to the presence of a NPXY internalization motif in the intracellular tail of the receptor. Endocytic uptake has been investigated thoroughly for mammalian LDLR (for reviews see Goldstein *et al.*, 1985; Jeon and Blacklow, 2005): endosomes undergo acidification, resulting in the release of lipoprotein from LDLR into the endosomal lumen and its subsequent degradation upon maturation of the endosome to a lysosome (Maxfield and McGraw, 2004). LDLR escapes lysosomal degradation as it is transferred to the endosomal recycling compartment, from which it is re-transported to the plasma membrane and becomes available again for another round of ligand-binding and internalization. Endocytic behavior has also been observed for an insect LDLR family member, the lipophorin receptor (LpR) of *L. migratoria*, that appeared to recycle its ligand, unlike the mammalian LDLR (Van Hoof *et al.*, 2002, 2003, 2005a, 2005b). This distinct behavior of *L. migratoria* LpR has been proposed to depend on cooperation between its ligand-binding and EGF precursor homology domains (Van Hoof *et al.*, 2005b). The physiological function of LpR-mediated ligand recycling following endocytosis remains to be established, as LpR is only temporarily expressed during early developmental stages in *L. migratoria* (Dantuma *et al.*, 1999; Van Hoof *et al.*, 2003), and lipophorin-mediated lipid transport in adult insects does not depend on endocytosis (Van Antwerpen *et al.*, 1988).

Scope of this thesis

As indicated above, the apparent divergence among animal lipoproteins may result from modifications in the structural components involved. From this perspective, this thesis focuses on the biogenesis of insect lipoprotein, and extends to the analysis of the diversity of lipoprotein precursors as well as their receptors. These studies were founded on a well established model insect that is known to be very active in lipid mobilization and utilization, the migratory locust *L. migratoria*. Lipoprotein biosynthesis in this animal is characterized by two major events: (1) the cleavage of the precursor apolipophorin, apoLp-II/I, into apoLp-I and –II (Weers *et al.*, 1992) and (2) lipidation of apoLp-II/I or its two cleavage products to a lipophorin of high buoyant density (Weers *et al.*, 1993).

The involvement of an insect furin in apoLp-II/I cleavage was investigated using fat body tissue *in vitro* as well as a novel recombinant insect expression system for truncated apoLp-II/I (**Chapter 2**). The results reveal the involvement of an insect furin in apoLp-II/I cleavage. Surprisingly, however, cleavage is not essential to the lipidation and secretion of lipophorin, as uncleaved apoLp-II/I was also secreted as a lipoprotein.

The mechanism that allows lipidation of the precursor apoLp-II/I or its two cleavage products to a high-density lipoprotein is unknown. Therefore, the involvement of two factors hypothesized to affect lipophorin biogenesis, apolipoprotein amphipathic

clusters and MTP, was studied (**Chapter 3**). Sequence analysis predicts that apoLp-II/I contains clusters of amphipathic α-helices and β-strands, organized along the protein as N-α_1-β-α_2-C. Expression of C-terminal truncation variants of *L. migratoria* apoLp-II/I in a novel recombinant lipoprotein expression system revealed the β cluster to be essential for HDLp formation. A *Drosophila melanogaster* MTP homolog stimulated recombinant apoLp-II/I lipidation and secretion, indicating that the lipoprotein assembly of insects, similar to that of mammals, depends on amphipathic clusters as well as MTP.

The evolutionary origin and diversity of the LLTP family and its defining domain, the LLT module, were explored in **Chapter 4**. Sequence databases were searched to categorize LLTP, resulting in the recognition of two novel LLTP in insects, in addition to Vtg, apoLp-II/I, and MTP. Phylogenetic analysis on conserved sequence blocks in the LLT module reveals the relatively close relationship of insect apoLp-II/I, decapod vitellogenin, and vertebrate apoB. Given its apparent absence in other taxa, the LLT module appears to be a metazoan innovation.

In **Chapter 5**, the non-recurring domains of LDLR family members across the animal kingdom were compared. Classification based on these domains recognizes insect lipophorin receptors, LpR, as a distinct subgroup of the LDLR family, unlike previous classification based on the number of cysteine-rich repeats in the ligand-binding domain. The distinct sequence motifs observed in the non-recurring domains of LpR may explain functional differences with mammalian LDLR family members.

The subsequent discussion (**Chapter 6**) summarizes and integrates the findings, and the molecular origin and function of specific insect lipoprotein properties are discussed.

Chapter 2

Biosynthesis and secretion of insect lipoprotein: involvement of furin in cleavage of the apoB homolog, apolipophorin-II/I

Marcel M.W. Smolenaars[1], Marcelle A.M. Kasperaitis[1], Paul E. Richardson[2], Kees W. Rodenburg[1], and Dick J. van der Horst[1]

[1] Biochemical Physiology, Department of Biology and Institute of Biomembranes, Utrecht University, The Netherlands

[2] Department of Chemistry, Coastal Carolina University, USA

J. Lipid Res. (2005) **46,** 412–421.

Abstract

The biosynthesis of neutral fat-transporting lipoproteins involves the lipidation of their non-exchangeable apolipoprotein. In contrast to its mammalian homolog apolipoprotein B, however, insect apolipophorin-II/I (apoLp-II/I) is cleaved post-translationally at a consensus substrate sequence for furin, resulting in the appearance of two apolipoproteins in insect lipoprotein. To characterize the cleavage process, a truncated cDNA encoding the N-terminal 38% of *Locusta migratoria* apoLp-II/I, including the cleavage site, was expressed in insect Sf9 cells. This resulted in the secretion of correctly processed apoLp-II and truncated apoLp-I. The cleavage could be impaired by the furin inhibitor decanoyl-Arg-Val-Lys-Arg-chloromethyl ketone (decRVKRcmk) as well as by mutagenesis of the consensus substrate sequence for furin, as indicated by the secretion of uncleaved apoLp-II/I-38. Treatment of *L. migratoria* fat body, the physiological site of lipoprotein biosynthesis, with decRVKRcmk similarly resulted in the secretion of uncleaved apoLp-II/I, which was integrated in lipoprotein particles of buoyant density identical to *wild-type* high-density lipophorin (HDLp). These results show that apoLp-II/I is post-translationally cleaved by an insect furin and that biosynthesis and secretion of HDLp can occur independent of this processing step. Structure modeling indicates that the cleavage of apoLp-II/I represents a molecular adaptation in homologous apolipoprotein structures. We propose that cleavage enables specific features of insect lipoproteins, such as low-density lipoprotein formation, endocytic recycling, and involvement in coagulation.

Introduction

Lipoproteins mediate most of the lipid transport in the circulation of animals. In mammals, a single protein component, the non-exchangeable apolipoprotein B (apoB), provides the structural basis for the biosynthesis of neutral fat-transporting lipoproteins (Mahmood Hussain *et al.*, 2001; Shelness and Sellers, 2001). Interestingly, the major lipoprotein of insects, lipophorin, contains two structural apolipoproteins, because the insect apoB homolog (Babin *et al.*, 1999; Mann *et al.*, 1999), the precursor apolipophorin-II/I (apoLp-II/I), is cleaved during lipoprotein biosynthesis (Van der Horst *et al.*, 1993; Weers *et al.*, 1993).

Cleavage of apoLp-II/I can be related to the activity of furin, a member of the proprotein convertase (PC) family of subtilisin-like serine endoproteases that is mainly active in the trans-Golgi network (Taylor *et al.*, 2003). The preferred consensus substrate sequence for furin, R-X-K/R-R, is present in all apoLp-II/I sequences characterized to date (Kutty *et al.*, 1996; Sundermeyer *et al.*, 1996; Bogerd *et al.*, 2000; Holt *et al.*, 2002). In accordance with the activity of furin, *Locusta migratoria* apoLp-II/I appears to be

cleaved immediately C-terminal of its furin substrate sequence, $RQKR^{720}$, as indicated by the N-terminal sequence of apoLp-I (Bogerd *et al.*, 2000).

The predicted furin cleavage site in each insect apoLp-II/I is located in the large lipid transfer (LLT) module, which constitutes the N-terminal region of apoLp-II/I that has sequence homology to that of apoB, microsomal triglyceride transfer protein (MTP), and vitellogenin (Babin *et al.*, 1999; Mann *et al.*, 1999). In apoB, this domain is essential for lipoprotein biosynthesis. The interaction between the LLT module of apoB and that of MTP enables the assembly of apoB-containing lipoproteins (Mahmood Hussain *et al.*, 2001; Shelness and Sellers, 2001). The homology between apoB and apoLp-II/I, as well as the presence of MTP in insects (Sellers *et al.*, 2003), suggest that the LLT module of apoLp-II/I enables lipoprotein biosynthesis in insects as well. Therefore, we hypothesized that the cleavage of apoLp-II/I in the LLT module functions in the biosynthesis and secretion of insect lipoprotein.

In the present report, we characterize the involvement of insect furin in the cleavage of *L. migratoria* apoLp-II/I and investigate the importance of this post-translational modification to insect lipoprotein biosynthesis and secretion. To this end, apoLp-II/I cleavage was investigated in a recombinant insect expression system for truncated apoLp-II/I as well as in the locust fat body, the insect tissue that expresses apoLp-II/I and secretes its cleavage products apoLp-I and -II together in the form of a high-density lipophorin (HDLp) particle (Weers *et al.*, 1992; Weers *et al.*, 1993). The results indicate that *L. migratoria* apoLp-II/I is cleaved by an insect furin. Uncleaved apoLp-II/I could be secreted and formed a high-density lipoprotein. We conclude that cleavage of apoLp-II/I is not required for the biosynthesis and secretion of mature lipophorin. Rather, this post-translational modification may represent a molecular adaptation enabling specific features of insect lipoproteins. Modeling of apoLp-II/I to the available homologous structures, the lipovitellin crystal structure and an apoB model, indicated the position of the cleavage site in the apoLp-II/I structure, and putative functions for apoLp-II/I cleavage are discussed accordingly.

Materials and Methods

Construction of the apoLp-II/I-38 expression plasmid. Standard cloning and sequencing procedures (Sambrook *et al.*, 1989; Bogerd *et al.*, 2000) were used to obtain an expression construct for *L. migratoria* apoLp-II/I. The apoLp-II/I-38 truncation variant was constructed by cloning the large *Eco*RI-*Xho*I cDNA fragment from pBK-CMV cDNA clone B20 (Bogerd *et al.*, 2000) into pGEM7-Zf (Promega). Subsequently, this fragment was transferred to plasmid pALTER-1 (Promega) using *Xba*I and *Sma*I digestion. The *Kpn*I-*Xho*I fragment of this plasmid, encoding apoLp-II/I, was cloned into the insect cell expression vector pIZ/V5-His (Invitrogen) using the same restriction sites, yielding

plasmid pIZ/V5-His-mod1. Subsequently, a short *Xho*I-*Sac*II primer fragment, containing a *Bgl*II site, was cloned into pIZ/V5-His-mod1, 3' of the cDNA, to allow fusion of the open reading frame of the apoLp-II/I cDNA sequence with the V5 epitope and 6xHis tag, yielding pIZ/V5-His-mod1* (oligonucleotide sequences of the *Xho*I-*Sac*II primer fragment are 5'-TCGAGGCATGCAGATCTGGCCCGC-3' and 5'-GGGCCAGATCTG-CATGCC-3'). In addition, the *Xho*I-*Cla*I apoLp-II/I cDNA fragment from pBK-CMV cDNA clone C5 (Bogerd *et al.*, 2000) was cloned into pGEM7-Zf, using the *Xho*I and *Cla*I restriction sites, to yield pGEM7-mod2. In a final cloning step, the *Xho*I-*BamH*I fragment from pGEM7-mod2 was cloned into pIZ/V5-His-mod1* using the unique *Xho*I and *Bgl*II sites of the latter plasmid. DNA sequencing of this plasmid confirmed that the resulting construct in pIZ/V5-His encodes the apoLp-II/I amino acid residues 1–1,287, joined in frame at the C-terminus with a vector-encoded V5 epitope as well as a 6xHis tag (Figure 1A). Excluding the 21 residue signal peptide (Bogerd *et al.*, 2000), this apoLp-II/I truncation variant corresponds to 38% of the total precursor protein and is referred to as apoLp-II/I-38.

Mutagenesis of apoLp-II/I-38. The basic amino acid residues at the P1, P2, and P4 positions (Schechter and Berger, 1967) within the predicted furin cleavage site RQKR^{720} in apoLp-II/I-38 were changed using the QuickChange Site-Directed Mutagenesis kit (Stratagene), according to the manufacturer's instructions. Mutants (Figure 1B) were created using the following forward primers, in combination with their complement reverse primers (codons to be modified are underlined; nucleotide positions to be mutated are shown in boldface): QQKR, 5'-ACAGAGAGATTTGAGAAAACATTC**CA**ACAGA-AACGATCGGTTTC-3'; DQKR, 5'-CAGAGAGATTTGAGAAAACATTC**GAC**CAGA-AACGATCGGTTTCAAAAGATGC-3'; RQAR, 5'-GAGATTTGAGAAAACATTCAG-ACAG**GC**ACGATCGGTTTCAAAAGATGCTG-3'; RQKQ, 5'-CACAGAGAGATTC-GAGAAGACATTCAGACAGAAA**CA**ATCGGTTTC-3'; QQAQ, 5'-CACAGAGAGA-TTCGAGAAGACATTC**CA**ACAG**GC**ACA**A**TCGGTTTC-3'. The forward and reverse primers were used in a PCR procedure of 18 cycles (1 min at 95ºC, 45 s at 58ºC, and 14 min at 68ºC) with 10 ng of the apoLp-II/I-38 construct as the DNA template. After *Dpn*I digestion, the reaction contents were used to transform into XL1-Blue cells (Stratagene). Single zeocin-resistant colonies were cultured and plasmids were isolated. The introduction of mutations in these plasmids was verified by DNA sequencing.

Stable expression of apoLp-II/I-38 constructs in the Sf9 cell line. *Spodoptera frugiperda* Sf9 cells were maintained in adherent culture in serum-free Insect-Xpress medium (Cambrex) in polystyrene flasks (Greiner) at 27ºC in a humidified atmosphere and passed twice each week. Transfections with the *wild-type* and mutant apoLp-II/I-38 constructs were performed using Cellfectin reagent (Invitrogen) according to the manufacturer's instructions. Stable transformants were selected using 400 μg/ml zeocin (Cayla) and

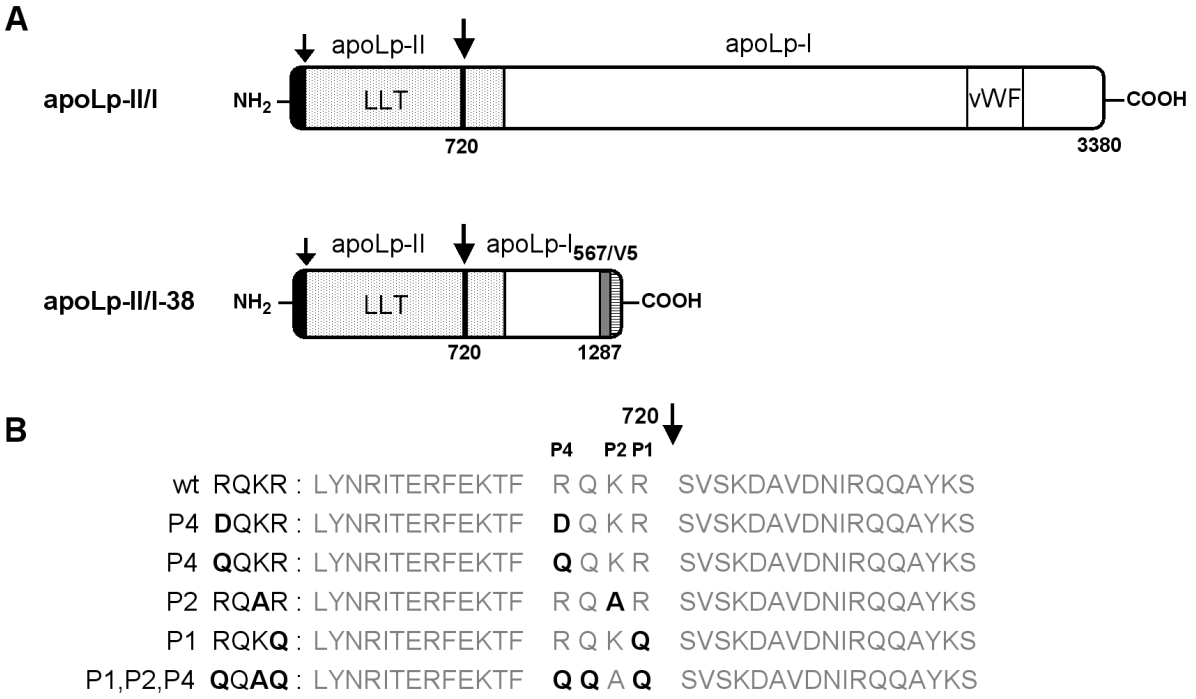


Figure 1. Expression constructs to characterize apolipophorin-II/I (apoLp-II/I) cleavage. A: The apoLp-II/I-38 cDNA construct encodes the N-terminal 38% of *L. migratoria* apoLp-II/I, including the cleavage site that is preceded by the consensus substrate sequence for furin at amino acids 717–720 and followed by 567 amino acids of apoLp-I, C-terminally fused to the V5 epitope and the 6xHis tag (referred to as apoLp-$I_{567/V5}$). LLT, large lipid transfer module; vWF, von Willebrand factor D module. The small arrows indicate the predicted signal peptide cleavage site, and the larger arrows indicate the putative site of cleavage by furin. **B**: Single and multiple mutations were introduced in the sequence of the apoLp-II/I-38 construct encoding the putative cleavage site at amino acids 717–720. P1 to P4 represent amino acid positions sequentially numbered starting N-terminal from the site of apoLp-II/I cleavage (Schechter and Berger, 1967). For each mutant, the residues 704–737 are shown, with mutated residues in boldface. The putative site of cleavage is indicated by the arrow.

maintained at 100 µg/ml zeocin. Experiments were performed in 25 cm^2 flasks with cells grown to 80% confluence. Incubations with the furin inhibitor decanoyl-Arg-Val-Lys-Arg-chloromethyl ketone (decRVKRcmk; Bachem) were performed by including either the inhibitor or its solvent methanol in fresh culture medium for 6 h. Culture media were collected and cleared from whole cells and cellular debris by centrifugation (1,000 *g*, 10 min; 22,000 *g*, 10 min, 4ºC), and analyzed for apoLp-II/I-38 and cleavage products by immunoblotting.

In vitro *incubation of* L. migratoria *fat body tissue.* *L. migratoria* were reared in cages under crowded conditions at 30ºC, 40% relative humidity, a light-dark cycle of 12 h/12 h, and a diet of reed grass supplemented with rolled oats. Fat bodies were dissected from day 9 adult males, separated in two halves along the length, and rinsed in saline buffer (10 mM HEPES, 150 mM NaCl, 10 mM KCl, 4 mM $CaCl_2$, 2 mM $MgCl_2$, pH 7.0).

Subsequent incubations were performed in a shaking water bath at 32ºC. After washing for 1 h at 32ºC in 4 ml of saline buffer, halves of six fat bodies were transferred to 250 µl of fresh saline buffer with 30 µCi of [^{35}S]Met/Cys (Promix; Amersham) supplemented with either 100 µM decRVKRcmk or an equivalent volume of its solvent methanol. Proteins in the incubation medium were analyzed for apoLp-II/I and cleavage products following density gradient analysis.

Density gradient analysis. The buoyant density of secreted apoLp-II/I and cleavage products was compared by subjecting the incubation media of decRVKRcmk-treated and control fat body tissue to KBr density gradient ultracentrifugation (Weers *et al.*, 1992). After gravimetric analysis of density, fractions were assessed for the presence of apoLp-II/I and cleavage products by SDS-PAGE (4–10% slab). Radiolabeled proteins in the gel were visualized by phosphorimaging on a Molecular Imager FX system with Quantity One software (Bio-Rad).

Immunoblot analysis of apoLp-II/I and cleavage products. Proteins were precipitated from incubation media by the addition of TCA to a concentration of 5%. Pellets were resuspended and heated (10 min at 95ºC) in Laemmli sample buffer that was modified by the addition of 5% (v/v) 1 M Tris to circumvent acidification by residual TCA. Protein samples were separated using SDS-PAGE (9% slab for transfected cell culture media; 4–10% slab for fat body tissue incubation media). Precision Plus Protein Dual Color Prestained Standards (Bio-Rad) were used as a protein molecular mass marker, and isolated *wild-type L. migratoria* HDLp (Weers *et al.*, 1993) was used as a positive control. Proteins were transferred to a polyvinylidene difluoride membrane (Millipore), and subsequent immunostaining was performed as described (Van Hoof *et al.*, 2002). Primary antibodies (1:10,000 dilution) were monoclonal α-V5 (Invitrogen), polyclonal α-II or α-I (Schulz *et al.*, 1987), or polyclonal α-IIC and α-IN (raised in rabbits against the peptides CKSLYNRITERFEKTFRQKR and SVSKDAVDNIRQQAYKSLLC, respectively, which correspond to the C-terminus of apoLp-II and the N-terminus of apoLp-I, excluding the terminal cysteines). Alkaline phosphatase-conjugated goat anti-rabbit or goat anti-mouse IgG (Jackson ImmunoResearch Laboratories) was used as a secondary antibody (1:10,000 dilution).

Modeling of* L. migratoria *apoLp-II/I. The *L. migratoria* apoLp-II/I sequence was used to identify homologous proteins and conserved domains using the web-based BLASTp program from the National Center for Biotechnology (Altschul *et al.*, 1990). The default settings were used for the BLASTp search. The most significant matches were to sequences from apoLp-II/I, apoB, and vitellogenin, including silver lamprey lipovitellin. Sequence homologies were located within the first 1,000 amino acid residues.

The alignment program CLUSTAL W (Thompson *et al.*, 1994) was used to align the homologous N-terminal regions of silver lamprey lipovitellin, human apoB-100, and *L. migratoria* apoLp-II/I. This alignment was used to generate the structural model of the first 1,009 amino acids (excluding the signal peptide) of locust apoLp-II/I using the program Modeller6 (Šali and Blundell, 1993), based on the crystal structure of silver lamprey lipovitellin (PDB 1LSH) (Thompson and Banaszak, 2002) and an all atom model for apoB (Richardson *et al.*, 2005). To eliminate steric problems and to optimize bond lengths and angles, the apoLp-II/I model was subjected to 250 steps of steepest descent energy minimization using the DISCOVER program package from INSIGHT2 (Accelerys, Inc.). A graphic representation of the model was generated by the Swiss-PdbViewer (Guex and Peitsch, 1997).

Results

Expression and proteolytic processing of apoLp-II/I-38 by Sf9 cells

To study the sequence characteristics of apoLp-II/I that enable post-translational cleavage, an apoLp-II/I truncation was recombinantly expressed in the insect Sf9 cell line, which expresses a furin homolog (Cieplik *et al.*, 1998). The construct used encodes the N-terminal 38% of apoLp-II/I (Figure 1A). This apoLp-II/I-38 polypeptide includes the signal peptide, the complete sequence for apoLp-II with the consensus substrate sequence for furin, $RQKR^{720}$, followed by 567 N-terminal residues from apoLp-I, fused to the V5 epitope and a 6xHis tag (referred to as apoLp-$I_{567/V5}$).

Stable transfection of Sf9 cells with this apoLp-II/I-38 construct resulted in the secretion of apoLp-II and apoLp-$I_{567/V5}$ cleavage products into the incubation medium, as demonstrated by immunoblot analysis (Figure 2A). Secretion of apoLp-$I_{567/V5}$ is indicated by the single reactive band obtained with antibodies directed against apoLp-I (α-IN and α-I; Figure 2A, lanes 4 and 5) as well as the V5 epitope (α-V5; Figure 2A, lane 6) at 66 kDa, similar to the estimated molecular mass of apoLp-$I_{567/V5}$. The weak immunoreactivity of polyclonal α-I (Figure 2A, lane 5) likely reflects the limited representation of its epitopes in apoLp-$I_{567/V5}$. Secretion of recombinant apoLp-II is indicated by immunoblotting with antibodies directed against apoLp-II (α-II and α-IIC; Figure 2A, lanes 2 and 3), which results in a single immunoreactive band at 72 kDa. This recombinant protein appears to migrate identically to the apoLp-II from purified *L. migratoria* HDLp (Figure 2A, lane 1 *vs.* lane 2). This similar size indicates that recombinant apoLp-II is similarly glycosylated as *wild-type* apoLp-II. Indeed, deglycosylation with endogycosidase H resulted in a single additional apoLp-II immunoreactive band with a decreased molecular mass of 3 kDa (Smolenaars *et al.*, 2005), which is similar to that reported for *wild-type* locust apoLp-II (Weers *et al.*, 1993).

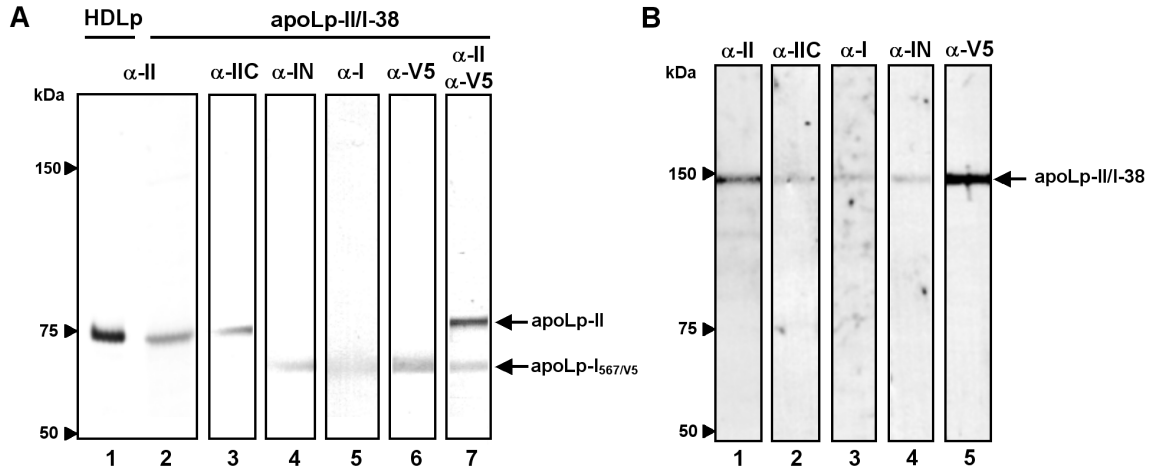


Figure 2. Recombinant expression and cleavage of apoLp-II/I-38. Incubation medium from stably transfected Sf9 cells was analyzed for the presence of the putative apoLp-II/I-38 cleavage products, apoLp-II and apoLp-$I_{567/V5}$, using antibodies directed against either apoLp-II (α-II and α-IIC) or apoLp-I (α-I, α-IN, and α-V5). **A**: Immunoblot analysis of media from Sf9 cells stably transfected with the apoLp-II/I-38 construct. **B**: Immunoblot analysis of medium from Sf9 cells stably transfected with the apoLp-II/I-38 construct in which the cleavage site was mutated from the *wild-type* sequence $RQKR^{720}$ to $QQAQ^{720}$. Molecular mass standards are indicated at the left. Arrows mark the positions of apoLp-II/I-38 and its cleavage products. For reference, lane numbers are indicated at the bottom.

Moreover, the similar migration behavior of recombinant and *wild-type* apoLp-II indicates that the cleavage of apoLp-II/I-38 proceeds identical to that in fat body (Weers *et al.*, 1993; Bogerd *et al.*, 2000). Expression of a construct expressing the N-terminal 33% of apoLp-II/I in Sf9 cells, and apoLp-II/I-38 in *Drosophila melanogaster* S2 cells, also resulted in the secretion of the two expected apoLp-II/I cleavage products (data not shown).

Recombinant apoLp-II/I-38 is cleaved at the same site as apoLp-II/I in *L. migratoria*. The α-IIC and α-IN antibodies, raised against synthetic oligopeptides of the sequences that flank the cleavage site N- and C-terminally, specifically recognize recombinant apoLp-II and apoLp-$I_{567/V5}$, respectively (Figure 2A, lanes 3 and 4). In addition, mutation of the three basic residues in the putative furin cleavage site (yielding the mutant designated $QQAQ^{720}$; Figure 1B) results in the secretion of uncleaved apoLp-II/I-38 (Figure 2B), as shown by the single immunoreactive band obtained with α-II, α-IIC, α-I, α-IN, and α-V5 antibodies at 145 kDa (Figure 2B, lanes 1–5, respectively), similar to the combined molecular mass of apoLp-II and apoLp-$I_{567/V5}$. The limited immunoreactivity of α-IIC antibody against mutant apoLp-II/I (Figure 2B, lane 2) likely relates to the introduced mutations. In further experiments, the degree of apoLp-II/I-38 cleavage was quantified using α-V5 antibody. Thus, Sf9 cells cleave recombinant apoLp-II/I-38 into apoLp-II and apoLp-$I_{567/V5}$ at the consensus substrate sequence for furin,

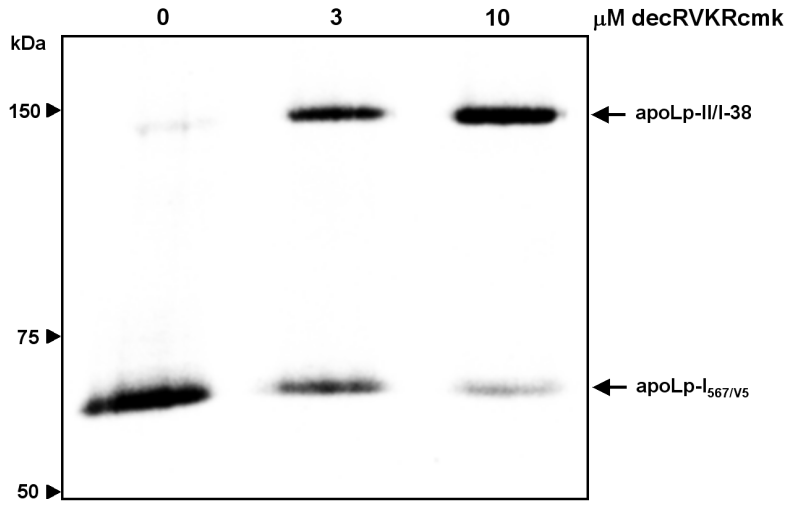


Figure 3. Effect of the proprotein convertase inhibitor decanoyl-Arg-Val-Lys-Arg-chloromethyl ketone (decRVKRcmk) on the secretion of recombinant apoLp-II/I-38. Sf9 cells stably transfected with apoLp-II/I-38 were incubated for 6 h with decRVKRcmk at concentrations of 0, 3, and 10 μM. Incubation media were analyzed for cleavage of apoLp-II/I-38 by immunoblot analysis with α-V5 antibody. Molecular mass standards are indicated at the left, and decRVKRcmk concentrations are shown above the blot. Arrows mark the positions of apoLp-II/I-38 and apoLp-$I_{567/V5}$.

similar to apoLp-II/I in locust fat body. These results validate the use of this expression system to characterize apoLp-II/I cleavage.

The secreted recombinant apoLp-II/I-38 cleavage products apoLp-II and apoLp-$I_{567/V5}$ have been characterized further for complex formation and lipidation. The coelution of apoLp-II with apoLp-$I_{567/V5}$ after affinity chromatography suggests that both cleavage products can form a complex (see Smolenaars *et al.*, 2005). Upon density gradient ultracentrifugation, the cleavage products apoLp-II and apoLp-$I_{567/V5}$ were found together in the fractions with densities between 1.20 and 1.25 g/ml, as was uncleaved mutant apoLp-II/I-38 ($QQAQ^{720}$) (see Smolenaars *et al.*, 2005; Chapter 3). These results indicate that the secreted recombinant apoLp-II/I-38 products are poorly lipidated. Therefore, the hypothesized role for apoLp-II/I cleavage in lipidation was assessed in *L. migratoria* fat body tissue, whereas the present expression system was used to characterize apoLp-II/I cleavage.

An inhibitor implicates PCs in the cleavage of recombinant apoLp-II/I

DecRVKRcmk is a modified tetrapeptide that irreversibly inhibits a wide range of PCs, including furin (Jean *et al.*, 1998) and Sfurin, the furin homolog characterized from the insect Sf9 cell line (Cieplik *et al.*, 1998). Incubation of Sf9 cells, transfected with the apoLp-II/I-38 construct, with micromolar concentrations of decRVKRcmk resulted in the secretion of uncleaved apoLp-II/I-38, whereas secretion of both apoLp-II and apoLp-

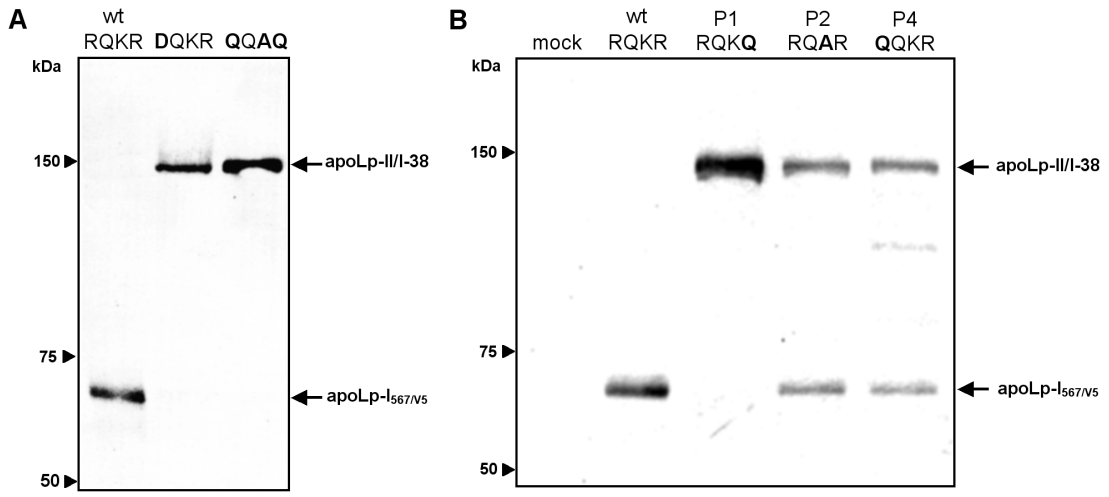


Figure 4. Cleavage of recombinant apoLp-II/I-38 upon modification of the consensus cleavage site for furin. Sf9 cells were stably transfected with *wild-type* (wt) and mutant apoLp-II/I-38 constructs (designated as in Figure 1B), and incubation media were analyzed for cleavage of apoLp-II/I-38 by immunoblotting with α-V5 antibody. **A**: Effect of a negatively charged amino acid at the P4 position, resulting in the mutant DQKR^{720}. *Wild-type* and QQAQ^{720} apoLp-II/I-38 transfectants were included to mark uncleaved and cleaved apoLp-II/I-38 products. **B**: Effect of the mutation of basic residues in the furin consensus substrate sequence to amino acids neutral of charge. Molecular mass standards are indicated at the left. Arrows mark the positions of apoLp-II/I-38 and apoLp-$I_{567/V5}$. Mutated residues are shown in boldface.

$I_{567/V5}$ was concomitantly and dose-dependently decreased, as judged by the immunoreactivity of α-V5 (Figure 3) and α-II (data not shown). Based on the amount of secreted protein, cleavage appeared to be prevented for up to 90% at a concentration of 10 μM decRVKRcmk. The sensitivity to decRVKRcmk indicates the involvement of a PC in the cleavage of apoLp-II/I.

A consensus substrate sequence for furin is required to efficiently process recombinant apoLp-II/I

To provide further evidence for the involvement of PCs in apoLp-II/I-38 cleavage, a negatively charged amino acid was introduced at position P4 (Figure 1B), which is expected to disturb binding by PCs via charge repulsion at the S4 binding pocket (Rehemtulla and Kaufman, 1992; Roebroek *et al.*, 1994; Henrich *et al.*, 2003). As demonstrated by immunoblot analysis of media from stably transfected Sf9 cells using α-V5 antibody, introduction of Asp at the P4 position (R to D; DQKR^{720}) completely prevented cleavage (Figure 4A), implicating the involvement of a PC in the cleavage of apoLp-II/I.

To further identify the sequence that directs the cleavage of apoLp-II/I, the basic residues in the furin consensus substrate sequence of apoLp-II/I-38 (RQKR^{720}) were mutated to neutral (uncharged) amino acids (Figure 1B). Mutagenesis at the P1 position

(R to Q; RQKQ^{720}) abrogated cleavage of apoLp-II/I-38 (Figure 4B). In contrast, mutation of the cleavage site residues at the P2 (K to A; RQAR^{720}) or P4 (R to Q; QQKR^{720}) position prevented cleavage only partially, reminiscent of the inefficient cleavage at imperfect cleavage sites described for mammalian furins (Rehemtulla and Kaufman, 1992; Rockwell *et al.*, 2002). Again, impairment of cleavage did not hamper the secretion of truncated apoLp-II/I by the recombinant expression system (Figure 4B). These results demonstrate that efficient apoLp-II/I cleavage in Sf9 cells requires the presence of the preferred consensus substrate sequence for furin, R-X-K/R-R, in apoLp-II/I. This indicates the involvement of an insect furin in apoLp-II/I cleavage.

Uncleaved native apoLp-II/I is secreted and forms a stable lipoprotein

To validate the involvement of an insect furin in lipophorin biosynthesis, and to establish the importance of cleavage for (apo)lipoprotein secretion in a physiological situation, fat body tissue from adult male locusts was treated *in vitro* with decRVKRcmk. Experiments showed the release of large amounts of HDLp that had adhered to fat body tissue. Therefore, [^{35}S]Met/Cys was included in incubation media to label newly biosynthesized proteins.

To identify secreted apoLp-II/I, incubation media were submitted to density gradient ultracentrifugation, and the resulting fractions were analyzed for radiolabeled proteins by phosphorimaging. Compared with control-treated fat body tissue, incubation with 100 μM decRVKRcmk resulted in the appearance of an additional radiolabeled band of high molecular mass and a concomitant decrease in the putative apoLp-I, in fractions with an average buoyant density of 1.12 ± 0.01 g/ml (Figure 5A), identical to the density previously reported for *wild-type* HDLp (Weers *et al.*, 1992). The identity of the radiolabeled proteins in these fractions as apoLp-II/I and apoLp-I was confirmed by immunoblotting with α-I (Figure 5B). Although apoLp-II could not readily be identified among other labeled proteins, immunoblotting demonstrated its presence in the same fractions as apoLp-II/I and apoLp-I (data not shown). The sensitivity of apoLp-II/I cleavage to decRVKRcmk incubation (60% with 100 μM inhibitor; Figure 5A) was reduced in the fat body, compared with the recombinant expression system (90% with 10 μM inhibitor; Figure 3). This difference may reflect reduced delivery of decRVKRcmk within the fat body, as the hydrophobic decanoyl group of this inhibitor may cause it to accumulate at the surface of the lipid droplets in this tissue. Thus, fat body tissue can secrete uncleaved apoLp-II/I with a buoyant density identical to *wild-type* HDLp. Gel filtration chromatography indicates that this secreted apoLp-II/I forms particles with a molecular size identical to *wild-type* HDLp (data not shown). Together, these results demonstrate that fat body can secrete uncleaved apoLp-II/I that has been integrated in a high-density lipoprotein similar to *wild-type* HDLp. Apparently, the biosynthesis as well as secretion of insect lipoprotein can occur independently of apoLp-II/I cleavage.

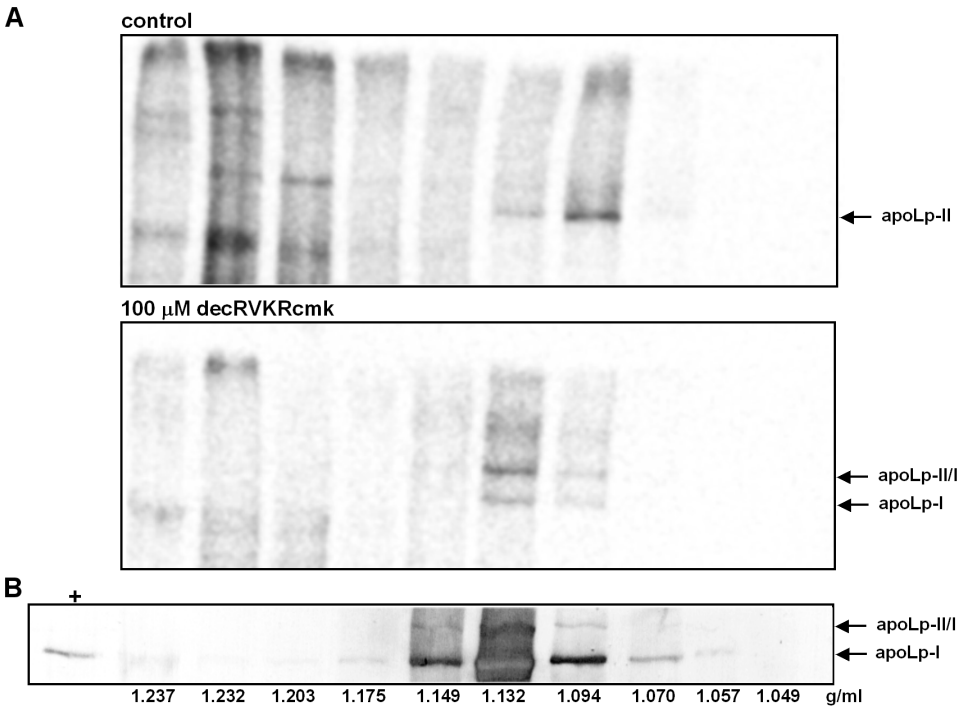


Figure 5. Buoyant density of apoLp-II/I secreted by fat body. A: Fat body tissue was labeled using [^{35}S]Met/Cys and incubated with 0 (control) or 100 μM decRVKRcmk. Incubation media were subsequently submitted to density gradient ultracentrifugation, fractions were analyzed by SDS-PAGE, and labeled proteins were visualized by PhosphorImager screen autoradiography. The signals of labeled proteins with molecular masses 150 kDa are shown for both the control and 100 M decRVKRcmk incubations, as indicated. **B**: Immunoblot detection of apoLp-II/I in ultracentrifugation fractions from decRVKRcmk-treated fat body tissue incubation media, using α-I antibody. Isolated *wild-type* high-density lipophorin was used as a positive control (as indicated by the plus sign). Arrows mark the positions of apoLp-I and apoLp-II/I. The density (g/ml) of each fraction is indicated at the bottom.

Homology modeling of apoLp-II/I indicates the position of the cleavage site

The ability of uncleaved apoLp-II/I to form a lipoprotein highlights the structural homology between insect and mammalian apoB-containing lipoproteins in the region of cleavage, but it also provokes questions regarding its function. To obtain a better understanding of the structural consequences of cleavage, *L. migratoria* apoLp-II/I was modeled to the available homologous structures, the lipovitellin crystal structure (Thompson and Banaszak, 2002) and an apoB model (Richardson *et al.*, 2005). The obtained apoLp-II/I model encompasses amino acids 22–1,030 and therefore the complete LLT module. This apoLp-II/I sequence can form a structure similar to that of the corresponding lipovitellin and apoB regions, as indicated by the recognition of three antiparallel β-sheets (βA, βB, and βC) and an extensive α-helical structure (Figure 6). The βA and βB sheets face each other to form a putative lipid-binding cavity and are partly surrounded by the α-helical region. The βC sheet is curved, resulting in a barrel-like

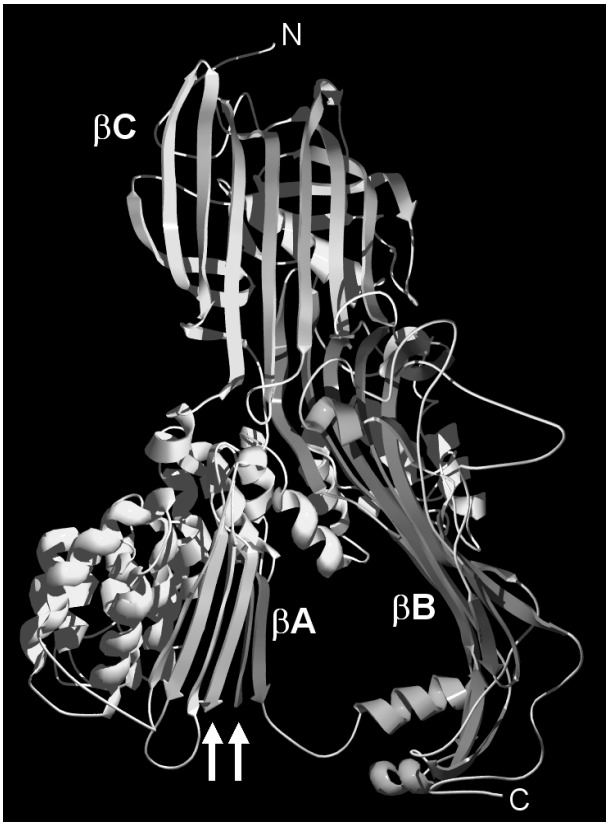


Figure 6. Model of *L. migratoria* apoLp-II/I. The model of locust apoLp-II/I includes amino acid residues 22–1,030 and was constructed based on sequence homology with silver lamprey lipovitellin and human apolipoprotein B, for which an all atom structure and a model are available, respectively (Thompson and Banaszak, 2002; Richardson *et al.*, 2005). The structures that are part of apoLp-I and apoLp-II after apoLp-II/I cleavage are marked in dark and light grey, respectively. The three β-sheets are indicated by βA, βB, and βC. N and C mark the amino- and carboxyl-terminal sides of the modeled region, respectively. The arrows indicate the β-strands at the base of the putative lipid-binding cavity that are connected by amino acid residues 669–748. ApoLp-II/I is cleaved within this region, between residues 720 and 721.

structure, which is situated above the putative lipid-binding cavity. The apoLp-II/I cleavage site, located between residues 720 and 721, was situated in an extended loop region (residues 669–748) that connects two β-strands of the βA sheet at the base of the putative lipid-binding cavity. This region could not be modeled confidently. The corresponding sequence was not resolved in the crystal structure of silver lamprey lipovitellin either (Thompson and Banaszak, 2002), which is suggestive of structural flexibility in this region of lipovitellin and possibly also in the region of apoLp-II/I cleavage. According to the present model, apoLp-II includes the βC sheet, the α-helical region, and most of the βA sheet, whereas apoLp-I consists of two distal β-strands from the βA sheet, the full βB sheet, and further unmodeled C-terminal sequences.

Discussion

It is well established that insect lipoproteins contain two apolipoproteins, apoLp-I and apoLp-II, that are derived from their common precursor protein apoLp-II/I (Weers *et al.*, 1993; Kutty *et al.*, 1996; Sundermeyer *et al.*, 1996; Bogerd *et al.*, 2000; Holt *et al.*, 2002). In the present study, we identified and characterized the cleavage site in *L. migratoria*

apoLp-II/I using an insect recombinant expression system of Sf9 cells and demonstrated that its cleavage likely involves an insect furin. As locust fat body can secrete uncleaved apoLp-II/I in the form of a lipoprotein particle with a buoyant density and molecular mass identical to *wild-type* HDLp, we conclude that cleavage of apoLp-II/I by insect furin is not required for the biosynthesis or the secretion of this insect lipoprotein.

The involvement of an insect furin in the cleavage of *L. migratoria* apoLp-II/I is indicated by three results. First, apoLp-II/I cleavage in both the Sf9 recombinant expression system and locust fat body can be inhibited by decRVKRcmk, which specifically inhibits PCs (Jean *et al.*, 1998), including Sfurin, the furin homolog characterized from the insect Sf9 cell line (Cieplik *et al.*, 1998). Second, cleavage is completely prevented by mutation of the basic amino acid residue (Arg) at the P4 position to a negatively charged amino acid residue (Asp). Based on the presence of a negatively charged residue at the S4 substrate binding pocket of PCs, this mutation is predicted to impair the substrate interaction via charge repulsion (Rehemtulla and Kaufman, 1992; Roebroek *et al.*, 1994; Henrich *et al.*, 2003). Third, complete proteolytic processing requires an intact furin consensus substrate sequence with basic amino acid residues at P1, P2, and P4. Mutation of any of these sites into amino acids neutral of charge results in the secretion of uncleaved apoLp-II/I, yet some cleavage does occur in the P2 (Lys→Ala) and P4 (Arg→Gln) mutants. Mammalian furin is known to cleave, with reduced activity, at sequences that partially match the preferred R-X-K/R-R substrate sequence (Rehemtulla and Kaufman, 1992; Rockwell *et al.*, 2002). Likewise, the partial cleavage in these P2 and P4 mutants may result from limited activity of furin at non-ideal sites. Together, these results indicate cleavage of apoLp-II/I by an insect furin.

Several furin homologs have been identified in insects. In *D. melanogaster*, cleavage activity and homology have been demonstrated for the dfurin1 and dfurin2 gene products (Roebroek *et al.*, 1991; Roebroek *et al.*, 1992; De Bie *et al.*, 1995). Furthermore, a furin was cloned from the insect Sf9 cell line (Cieplik *et al.*, 1998). Moreover, a PC with furin-like activity is present in the fat body of *Aedes aegypti* (Chen and Raikhel, 1996). However, no putative furin has yet been identified in *L. migratoria* fat body.

In transfected Sf9 cells as well as locust fat body, uncleaved apoLp-II/I could be secreted in similar amounts as its cleavage products. The recombinant apoLp-II/I-38 cleavage products apoLp-II and apoLp-$I_{567/V5}$ appeared to form a complex, yet were not integrated in a high-density lipoprotein, in contrast to apoLp-II/I products secreted by locust fat body, as they were recovered in the very high-density range upon density gradient ultracentrifugation (see Smolenaars *et al.*, 2005; Chapter 3). The poor lipidation of the apoLp-II/I-38 truncation products is in accordance with the decreased lipidation observed for apoB truncations (Shelness and Sellers, 2001; Richardson *et al.*, 2005) but may also reflect the absence of co-factors for lipidation in the present expression system. Therefore, the possible role of cleavage in the lipidation and secretion of insect lipoprotein was investigated in locust fat body. Here, decRVKRcmk treatment resulted in the

secretion of uncleaved apoLp-II/I that formed a stable high-density lipoprotein, as indicated by its density and molecular mass identical to *wild-type* HDLp. Consequently, the uncleaved precursor apoLp-II/I can function as a single apolipoprotein in the formation of lipoprotein, like its mammalian homolog apoB. Thus, apoLp-II/I can be lipidated to form a high-density lipoprotein, irrespective of its cleavage.

The lipidation of apoB starts co-translationally in the rough endoplasmic reticulum and is completed post-translationally in the smooth endoplasmic reticulum and/or cis-Golgi network, possibly by fusion with an intralumenal lipid droplet (Mahmood Hussain *et al.*, 2001; Shelness and Sellers, 2001). The lipidation process starts with the lipidation of the lipid-binding cavity in the apoB LLT module and requires the lipid transfer activity of MTP (Mahmood Hussain *et al.*, 2001; Shelness and Sellers, 2001). Based on the homology between apoB and apoLp-II/I (Babin *et al.*, 1999; Mann *et al.*, 1999) as well as the discovery of an insect MTP (Sellers *et al.*, 2003), insect lipoprotein assembly may also occur early in the secretory pathway. However, cleavage by furin homologs is performed late in the secretory pathway, mainly in the trans-Golgi network (Molloy *et al.*, 1999). Therefore, we propose that insect lipoprotein biosynthesis by the fat body involves lipidation of apoLp-II/I to a lipoprotein first, followed by cleavage of apoLp-II/I into apoLp-I and -II. The occurrence of cleavage before any lipidation might result in the parting of apoLp-I and apoLp-II, and hence impairment of lipoprotein biosynthesis. The uncleaved LLT module in apoLp-II/I may be essential to enable the first steps in lipidation, as in apoB.

Homology modeling shows that the LLT module of apoLp-II/I can form a putative lipid-binding cavity. Remarkably, the cleavage site is located in an unresolved loop region between two β-strands in the βA sheet at the base of this cavity (Figure 6). In apoB, the corresponding region is proposed to function in early lipidation events, as it may form a hairpin structure that temporarily closes the basal opening of the lipid-binding cavity by connecting to the βB sheet via salt bridges (Richardson *et al.*, 2005). When the lipid-binding cavity of apoB reaches a certain lipid content during lipoprotein assembly, these salt bridges are proposed to dissociate. This would allow for widening of the V-shaped lipid-binding cavity formed by the βA and βB sheets and the progression of apoB lipidation (Manchekar *et al.*, 2004; Richardson *et al.*, 2005). The cleavage of apoLp-II/I might represent an alternative structural solution to unlock the lipid-binding cavity and increase the flexibility of the non-exchangeable apolipoprotein, hence allowing for further lipidation of the lipoprotein particle. Interestingly, insects can lipidate circulating HDLp to a low-density lipophorin (LDLp) during conditions that require enhanced lipid transport (*e.g.*, long-term flight and vitellogenesis) (Kawooya and Law, 1988; Ryan and Van der Horst, 2000). Therefore, apoLp-II/I cleavage may function to enable the formation of LDLp from HDLp.

The apparent conservation of apoLp-II/I cleavage in all insects characterized to date reveals the importance of this processing step. Besides LDLp formation, however,

apoLp-II/I cleavage may enable other unique insect lipoprotein characteristics, such as the ability of lipophorin to function as a reusable lipid transporter (Ryan and Van der Horst, 2000) and to be recycled after endocytic uptake by an insect member of the low-density lipoprotein receptor (LDLR) family (Van Hoof *et al.*, 2002). Vitellogenin, another ligand of this receptor family, is homologous to apoB and apoLp-II/I, and is also cleaved at a furin consensus substrate sequence in the LLT module during biosynthesis in most insect species, but not vertebrates (Sappington and Raikhel, 1998a). Lipophorin, vitellogenin, and LDLR family members are involved in insect vitellogenesis, the transfer of nutrients to the developing oocyte (Sappington and Raikhel, 1998a), and perhaps the post-translational cleavage of apoLp-II/I and vitellogenin facilitates this process (*e.g.*, by enabling receptor binding). In addition, apoLp-II/I cleavage might enable a function for lipophorin in coagulation. Lipophorin has been implicated in this process by its abundant presence in clots (Brehélin, 1979; Gellissen, 1983; Barwig, 1985; Li *et al.*, 2002), capacity to aggregate (Chino *et al.*, 1987; Van Antwerpen, 1989), and interactions with other insect hemostasis factors (Theopold and Schmidt, 1997). Moreover, the disappearance of apoLp-I but not apoLp-II from plasma during coagulation (Duvic and Brehélin, 1998) suggests that apoLp-II/I cleavage enables distinct roles for apoLp-I and apoLp-II in coagulation.

Cleavage of the insect apoB homolog apoLp-II/I appears to be a molecular adaptation within homologous structures and may have a significant impact on insect lipoprotein function. The possibility of obtaining apoLp-II/I lipoprotein from fat body using a furin inhibitor may aid in establishing the physiological role of apoLp-II/I cleavage. Furthermore, the recombinant expression of truncated apoLp-II/I may be used to explore aspects of insect lipoprotein assembly and structure, such as lipid binding regions in apoLp-I and -II, the role of protein-protein and lipid-protein interactions in apoLp-I:apoLp-II complex formation and lipoprotein solubility (Kawooya *et al.*, 1989), as well as the co-factors involved in insect lipoprotein biosynthesis. For example, apoB truncations with a length similar to apoLp-II/I-38 can form high-density lipoprotein particles upon recombinant expression, but only when cells are supplied with exogenous lipids and (co)express MTP (Sellers *et al.*, 2003; Shelness *et al.*, 2003). Likewise, the present recombinant expression system may be used to investigate putative roles for lipid availability and insect MTP homologs in the biosynthesis of insect lipoprotein.

Acknowledgements

The authors thank Monique van Oers and Just Vlak (Department of Virology, Wageningen University, Wageningen, The Netherlands) for providing Sf9 cells, Karine Valentijn and Yvonne Derks for providing the α-IIC and α-IN antibodies as well as purified insect lipoprotein, Jan van Doorn for technical assistance, and Jana Kerver for help in obtaining the apoLp-II/I construct. The authors thank Jere Segrest (Department of Medicine, University of Alabama at Birmingham Medical Center) for his initiating role in the modeling work.

Chapter 3

Insect lipoprotein biogenesis depends on an amphipathic β cluster in apolipophorin-II/I and is stimulated by microsomal triglyceride transfer protein

Marcel M.W. Smolenaars, Antoine de Morrée, Jana Kerver, Dick J. van der Horst, and Kees W. Rodenburg

Biochemical Physiology, Department of Biology and Institute of Biomembranes, Utrecht University, The Netherlands

to be submitted

Abstract

Lipoproteins transport lipids in the circulation of an evolutionary wide diversity of animals. The pathway for lipoprotein biogenesis has been uncovered only in mammals, where apolipoprotein B (apoB) acquires lipids via the assistance of microsomal triglyceride transfer protein (MTP) and binds them by means of amphipathic protein structures. To investigate whether this is a common mechanism for lipoprotein biogenesis, we studied the structural elements involved in the assembly of lipophorin, the insect lipoprotein. LOCATE sequence analysis predicted that the insect lipoprotein precursor apolipophorin-II/I (apoLp-II/I) contains clusters of amphipathic α-helices and β-strands, organized along the protein as N-α_1-β-α_2-C, corresponding to a truncated apoB form. Recombinant expression of a series of C-terminal truncations of *Locusta migratoria* apoLp-II/I in an insect expression system revealed that formation of a buoyant high-density lipoprotein requires the amphipathic β cluster. Co-expression of apoLp-II/I with an insect MTP homolog impacted insect lipoprotein biogenesis quantitatively as well as qualitatively, as the secretion of apoLp-II/I proteins was increased several-fold while the buoyant density of the secreted lipoprotein concomitantly decreased, indicative of augmented lipidation. In addition, we observed that insect MTP is secreted, suggesting a possible extracellular function in insect lipoprotein assembly. Based on the above findings we propose that, despite specific modifications, the assembly of lipoproteins involves MTP as well as amphipathic structures in the apolipoprotein carrier, both in mammals and insects. Thus, lipoprotein biogenesis in animals appears to rely on structural elements that are of early metazoan origin.

Introduction

Lipoproteins transport lipids in the circulation of vertebrates as well as invertebrates. In mammals, the biosynthesis of these macromolecular protein:lipid complexes involves the intracellular transfer and subsequent stabilization of lipids onto a single protein component, the non-exchangeable apolipoprotein B (apoB) (for reviews see Shelness and Sellers, 2001; Mahmood Hussain *et al.*, 2001; Olofsson and Borén, 2005). To attain a comparative viewpoint on apoB lipoprotein assembly we investigated the biosynthesis of an invertebrate lipoprotein, *i.e.* lipophorin, the insect lipoprotein.

The structural protein of lipophorin is a homolog of apoB (Babin *et al.*, 1999). Remarkably, this apolipophorin-II/I (apoLp-II/I) is cleaved during lipoprotein biosynthesis, resulting in the presence of two proteins in insect lipoprotein, apoLp-I and apoLp-II (Weers *et al.*, 1993; Babin *et al.*, 1999; Chapter 2). Cleavage is not essential for lipoprotein biogenesis, as uncleaved apoLp-II/I also formed a lipoprotein (Chapter 2).

Lipophorin is produced in cells of the fat body and is secreted with a buoyant density in the high-density or very high-density lipoprotein range, dependent on the species (Prasad *et al.*, 1986; Venkatesh *et al.*, 1987; Capurro and De Bianchi, 1990; Weers *et al.*, 1992; Van Heusden *et al.*, 1998). Accordingly, lipophorin contains far less associated lipids than apoB lipoproteins that are secreted as very low-density lipoproteins. Moreover, the major neutral lipid in the lipophorin of most insects is diacylglycerol (DAG), rather than the triacylglycerol (TAG) that dominates the lipid content of newly biosynthesized apoB lipoproteins (Weers *et al.*, 1992; Soulages and Wells, 1994). It remains to be established whether lipidation of apoLp-II/I (cleavage products) is completed intracellularly, as for apoB (for reviews see Shelness and Sellers, 2001; Mahmood Hussain *et al.*, 2001; Borén and Olofsson, 2005). Lipid transfer particle (LTP), an insect-specific heterotrimer that can mediate the transfer of lipids between lipoproteins and tissues (for review see Ryan and Van der Horst, 2000), has been suggested to complete lipidation at the fat body plasma membrane (Capurro and De Bianchi, 1990). Therefore, insect lipoprotein biosynthesis differs from mammalian apoB lipoprotein formation in the amount and type of lipid acquired, cleavage of the structural apolipoprotein, and possibly the localization of lipidation.

Unlike the biosynthesis of lipophorin, the molecular mechanism of apoB lipoprotein biogenesis has been studied extensively (for reviews see Shelness and Sellers, 2001; Mahmood Hussain *et al.*, 2001; Borén and Olofsson, 2005). ApoB lipoprotein biogenesis starts upon translation of apoB mRNA and translocation of the nascent polypeptide into the ER. Following translocation, the N-terminal ~900 amino acids of apoB fold into the large lipid transfer (LLT) module that interacts with microsomal triglyceride transfer protein (MTP). This interaction is required for the deposition of lipids in the lipid-binding cavity of apoB's LLT module. Accordingly, MTP prevents misfolding of the more C-terminal amphipathic apoB sequences and the resulting intracellular retention and degradation of apoB. Following apoB translocation, the lipoprotein particle continues along the secretory pathway and acquires the bulk of its neutral lipids (TAG), putatively by fusion with an intralumenal neutral lipid droplet.

ApoB can bind hundreds of lipid molecules by amphipathic structures that ultimately envelope and stabilize a core of lipids, thus shielding hydrophobic lipids from the hydrophilic environment (for reviews see Hevonoja *et al.*, 2000; Segrest *et al.*, 2001). Prediction of amphipathic α-helices and amphipathic β-strands in apoB by the program LOCATE suggested the presence of five amphipathic clusters, enriched in either of secondary structure elements, organized along the apoB polypeptide as N-α_1-β_1-α_2-β_2-α_3-C (Segrest *et al.*, 1994, 1998). The α_1 cluster and the N-terminal part of the β_1 cluster constitute the LLT module. The C-terminal β_1-α_2-β_2-α_3 clusters stabilize expansion of the initial lipid core in the LLT module and actually possess most lipid-binding capacity (for review see Segrest *et al.*, 2001). The β_1 cluster appears to be especially important to

buoyant lipoprotein formation, and the mammalian intestine in fact produces lipoproteins of very low buoyant density from C-terminally truncated apoB that contains little more than the α_1 and β_1 clusters, due to editing of the apoB RNA (for review see Davidson and Shelness, 2000). Thus, mammalian lipoprotein assembly is enabled by the molecular architecture of apoB as well as MTP.

Despite apparent differences in the biogenesis of insect apoLp-II/I and mammalian apoB lipoproteins, the presence of an LLT module in apoLp-II/I (Babin *et al.*, 1999; Chapter 2) and MTP homologs in insect genomes (Sellers *et al.*, 2003) suggest that lipoprotein biosynthesis in insects may be dictated by similar structural elements as in mammals. Therefore we studied the involvement of two major elements of mammalian lipoprotein biogenesis, amphipathic clusters in the structural apolipoprotein as well as MTP, in the assembly of *Locusta migratoria* apoLp-II/I into a high-density lipoprotein (HDLp), the mature form of lipophorin that enables the circulatory transport of lipids (fo review see Van der Horst *et al.*, 2001). Secondary structure prediction as well as recombinant expression of apoLp-II/I and *Drosophila melanogaster* MTP (dMTP) in insect cells provide direct evidence for a major role of apoLp-II/I amphipathic clusters and MTP in the biosynthesis of insect lipoprotein. Consequently, mechanisms of lipoprotein assembly in vertebrates and invertebrates appear to have a common evolutionary origin.

Materials and Methods

LOCATE analysis. LOCATE (Segrest *et al.*, 1994, 1998, 1999, 2001) was used to predict amphipathic α-helices and β-strands with high lipid affinity in apoLp-II/I and apoB sequences. The following sequences were downloaded from GenBank in FASTA format and analyzed: apoLp-II/I from *L. migratoria* (accession CAB51918.1), *D. melanogaster* (acc. AAC47284.1) and *Manduca sexta* (acc. AAB53254.1), and apoB from *Homo sapiens* (acc. P04114), *Gallus gallus* (acc. XP_419979.1) and *Danio rerio* (acc. XP_694827.1). Amphipathic α-helices and β-strands in these sequences were predicted by LOCATE. Default parameter settings were used, except the exhaustive search option was activated, as well as the snorkel option for amphipathic α-helix prediction. Lipid affinities were calculated using the Λ_3 algorithm (Segrest *et al.*, 1998).

Construction of the dMTP expression plasmid. A cDNA clone (SD01502) encoding dMTP was obtained from the Berkeley Drosophila Genome Project. To express dMTP in insect Sf9 cells, its complete coding sequence was transferred into the pIB/V5-His vector (Invitrogen). To this end, the dMTP coding sequence was PCR amplified (forward primer 5'-CGCGGCCGATATCATGGAGAACAAAAATAAGAAGTGCCTG-3', reverse primer 5'-GCAGGGCTCGAGTACAAGTCCCTAAAAATGAGGTTGCAC-3'). The

PCR product was purified and ligated into the pGEM-T vector (ProMega), and subsequently transferred into the pIB/V5-His vector using *Eco*RV/*Xho*I double digestion, in frame with the vector-encoded V5/His epitope tag. To prevent inclusion of possible PCR-introduced mutations, an *Eco*RI/*Afe*I dMTP fragment was transferred from the original cDNA clone into the dMTP:pIB/V5-His construct. Subsequent DNA sequencing verified that the dMTP cDNA sequence in the obtained expression construct was correct.

Construction of truncated and full-length apoLp-II/I expression plasmids. Subclones of the *L. migratoria* apoLp-II/I cDNA (Bogerd *et al.*, 2000) were used to step-wise extend a previously established apoLp-II/I-38 construct, consisting of the pIZ/V5-His expression vector encoding 38% (1290 aa) of the complete apoLp-II/I polypeptide (Chapter 2). A fragment corresponding to apoLp-II/I mRNA nucleotides 3387 to 6021 was amplified from one of the apoLp-II/I cDNA subclones by PCR (forward primer 5'-CAGAAGGAATCACTGTTGATGG-3'; reverse primer 5'-TCTAGATGGCCATTTCC-ACTGTATG-3') using *Pfu*Turbo DNA polymerase (Stratagene). Following Super*Taq* DNA polymerase (HT Biotechnologies Ltd.) mediated tailing with a terminal adenosine nucleotide, this fragment was cloned into pGEM-T. The transfer of a *Pac*I/*Sac*II fragment from the selected clone into the apoLp-II/I-38 clone (in the pIZ/V5-His vector) resulted in the apoLp-II/I-59 expression vector that encodes the N-terminal 2004 amino acids (aa) of apoLp-II/I, in frame with the V5-epitope and 6xHis-tag. The apoLp-II/I fragment in pGEM-T was subsequently extended up to apoLp-II/I mRNA nucleotide 8505 by the transfer of a *Bae*I-*Xba*I digested fragment from an apoLp-II/I cDNA subclone. The subsequent transfer of a *Pac*I/*Sac*II fragment from the resulting pGEM-T contruct into the apoLp-II/I-38 clone, resulted in the expression vector apoLp-II/I-84 that encodes the N-terminal 2832 aa of apoLp-II/I, in frame with the V5-epitope and 6xHis-tag. The subsequent cloning of a *BstE*II-digested PCR amplified fragment (forward primer 5'-CTGGTGGTGACCATACAAAGG-3'; reverse primer 5'-TCTAGACCTCCTTTAACA-CCCTTAG-3') into *BstE*II-*Xba*I digested apoLp-II/I-84 construct, resulted in the apoLp-II/I-100 construct that encodes full-length 3380 aa *L. migratoria* apoLp-II/I with a C-terminal V5-epitope and 6xHis-tag. The sequence of obtained constructs was verified for out-of-frame errors by DNA sequencing of both sense and anti-sense strands. Several non-synonymous and synonymous nucleotide differences as well as three nearly adjacent nucleotide deletions and one short 16-nucleotide frame-shifted segment, were observed as compared to the published sequence (Bogerd *et al.*, 2000). These differences likely reflect inaccuracies in the published sequence, and are to be deposited at GenBank as an update to the *L. migratoria* apoLp-II/I nucleotide and protein sequences.

Transient transfection and incubation of Sf9 cells. *Spodoptera frugiperda* Sf9 cells were maintained in adherent culture in serum-free Insect-Xpress medium (Cambrex) in polystyrene flasks (Greiner) at 27°C, and passed twice a week. Transfections were

performed in 6-well plates with cells grown to 80% confluency. Transfection mixture was prepared by adding DNA and 6 µg polyethylenimine (Polysciences Europe) to 50 µl culture medium. After 15 min incubation at room temperature, the transfection mix was added to the wells (put on 600 µl culture medium). Cells were incubated with DNA for 5 h, and subsequently washed with culture medium. The amount of DNA transfected was increased according to construct size, using 0.75, 0.75, 1, 1.2, and 1.5 µg per well for the dMTP, apoLp-II/I-38, -59, -84 and –100 constructs, respectively. For co-expression experiments, transfection mixtures were supplemented with either 0.75 µg dMTP construct or 0.3 µg control (pIB/V5-His) vector. For determinations of buoyant density and antibody reactivity, transfections were performed in 75 cm^2 tissue culture flasks, with reaction volumes for transfection scaled up according to surface increase. Incubations were performed 40 h post-transfection. Cells were rinsed twice with culture medium and incubated for 7 h in PBS (pH 7.4) supplemented with 0.75 mM $CaCl_2$, 0.75 mM $MgCl_2$ and 1 g/l β-D-glucose. Incubation media were subsequently sampled as described previously (Chapter 2).

Density gradient analysis. The buoyant density of recombinant apoLp-II/I proteins as well as dMTP was analyzed by subjecting incubation media from transfected cells to KBr density gradient ultracentrifugation (Weers *et al.*, 1992). Resulting gradients were divided into 11 fractions, collected from bottom to top. Following gravitometric analysis of density, fractions were TCA precipitated and assessed for the presence of apoLp-II/I proteins and dMTP by immunoblotting.

Gel filtration chromatography. Incubation media from apoLp-II/I transfected cells were concentrated 8-fold and submitted to Superose 12 and Superose 6 gel filtration chromatography. Proteins were eluted (130 mM NaCl, 5 mM KCl, 1.9 mM NaH_2PO_4, 1.7 mM K_2HPO_4, 5 mM EDTA, 0.2% NaN_3, pH 7.5) in 0.65 ml fractions that were TCA precipitated and analyzed for recombinant apoLp-II/I and its cleavage products by immunoblotting. Molecular size was calculated using a standard curve obtained from the elution peaks of the marker proteins bovine serum albumin (66 kDa), aldolase (158 kDa), katalase (240 kDa), ferritin (450 kDa) and thyroglobulin (669 kDa) (Sigma or Boehringer Mannheim).

Immunoblot analysis of apoLp-II/I and cleavage products. TCA-precipitated proteins from incubation media were analyzed for the presence of recombinant apoLp-I, apoLp-II, uncleaved precursor apoLp-II/I and dMTP by SDS-PAGE (5-10% slab) and subsequent immunoblot analysis (Chapter 2). Primary antibodies (dilution 1:10,000) were either monoclonal α-V5 (Invitrogen), or the polyclonals α-II and α-I that are directed against apoLp-II and apoLp-I, respectively (Schulz *et al.*, 1987).

Results

Amphipathic clusters in apoLp-II/I

The program LOCATE (Segrest *et al.*, 1994, 1998, 1999, 2001), used to predict amphipathic secondary structures (α-helices and β-strands) of high lipid affinity in available insect apoLp-II/I sequences, indicates the presence of two clusters of amphipathic α-helices of relatively high calculated lipid affinity in the *Locusta migratoria* apoLp-II/I, located between aa 275 to 750 and 2400 to 2750 (Figure 1A), named α_1 and α_2, respectively. These regions correspond to part of the LLT module and a less-conserved region just N-terminal to the von Willebrand Factor D (vWF-D) module (aa 2815 to 3004; Babin *et al.*, 1999), respectively. Compared to the α_1 cluster, the α_2 cluster contains a higher density of predicted amphipathic α-helices, that also have a higher calculated lipid affinity. Most amphipathic β-strands are between apoLp-II/I aa 750 and 2250, with a concentration of amphipathic β-strands of particularly high calculated lipid affinity between aa 800 and 950. The C-terminal part of apoLp-II/I (aa > 2750) is not clearly enriched in either amphipathic α-helices or amphipathic β-strands (Figure 1A).

A similar clustering of amphipathic α-helices and β-strands is predicted for other characterized insect apoLp-II/I (Figure 1C), that also appear to contain a region enriched in amphipathic β-strands in between two clusters of amphipathic α-helices, and a C-terminal part that is not particularly enriched in either of amphipathic structures. This organization of amphipathic clusters in insect apoLp-II/I may be summarized as N-α_1-β-α_2-C. Compared to LOCATE analysis of vertebrate apoB (Figure 1B, C), insect apoLp-II/I is predicted to contain three amphipathic clusters rather than five. These three clusters may correspond to the three N-terminal clusters of vertebrate apoB (N-α_1-β_1-α_2-β_2-α_3-C), although several differences were found. In all analyzed apoLp-II/I, the β cluster occupies at least 200 residues more than either of the β clusters of analyzed vertebrate apoB (Figure 1C). In addition, the β cluster of apoLp-II/I is predicted to contain amphipathic β-strands of particularly high calculated lipid affinity from aa 800 to 950, at its N-terminal boundary, unlike the β_1 and β_2 clusters in apoB (Figure 1B). These sequences correspond to the βB sheet that constitutes one side of the lipid-binding cavity of apoLp-II/I's LLT module (Chapter 2). The α_2 cluster of apoLp-II/I is smaller than the α_2 and α_3 clusters in vertebrate apoB, comprising ~300 rather than ~500 residues. Thus, prediction of amphipathic α-helices or amphipathic β-strands by LOCATE indicates the presence of amphipathic clusters enriched in either of amphipathic structures in apoLp-II/I, in an organization along the protein corresponding to N-α_1-β-α_2-C.

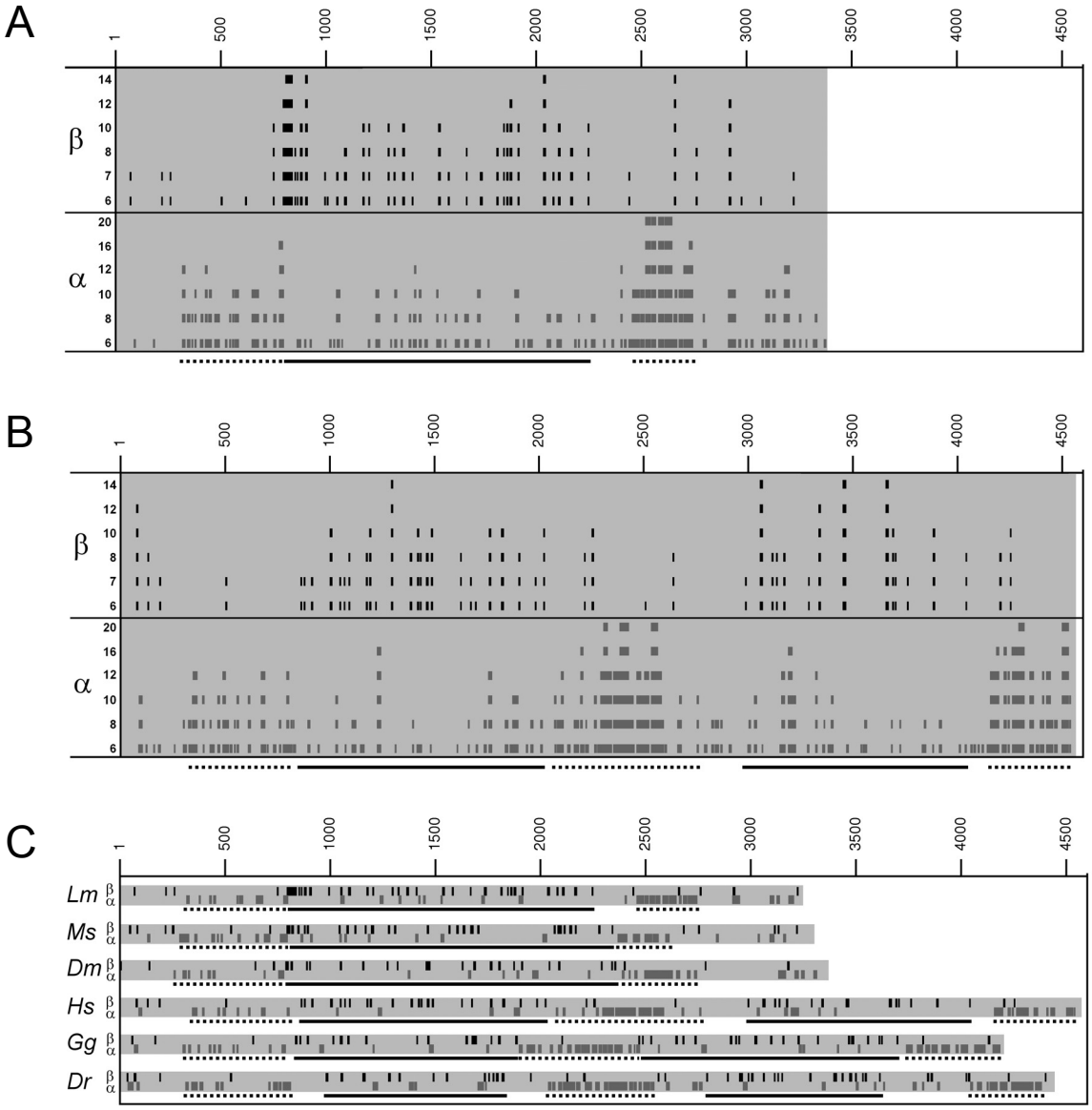


Figure 1. Clusters of amphipathic secondary structure in apoLp-II/I. Amphipathic α-helices and β-strands in apoLp-II/I and apoB were predicted using LOCATE. Predicted α-helical or β-strand amphipathic stretches are indicated by boxes at their predicted amino acid positions (horizontal axis), provided that their calculated lipid affinity is above the chosen threshold for calculated lipid affinity (vertical axis). Upper panels show predicted amphipathic β-strands and lower panels predicted amphipathic α-helices. Solid and dotted black lines below the panel of each LLTP indicate the designated clusters enriched in either amphipathic α-helices or amphipathic β-strands, respectively. **A**: LOCATE analysis of *L. migratoria* apoLp-II/I with threshold for calculated lipid affinity increasing from 6 to 20 and 6 to 14 kcal/mol for amphipathic α-helices and β-strands, respectively. **B**: LOCATE analysis of *H. sapiens* apoB with threshold for calculated lipid affinity increasing from 6 to 20 and 6 to 14 kcal/mol for amphipathic α-helices and β-strands, respectively. **C**: Amphipathic α-helices and β-strands predicted within apoLp-II/I from *L. migratoria*, *D. melanogaster* and *M. sexta* as well as apoB from *H. sapiens* (human), *G. gallus* (chicken), and *D. rerio* (zebrafish) at a calculated lipid affinity threshold of 10 and 7 kcal/mol, respectively.

Expression of apoLp-II/I truncations

To assess the role of the identified apoLp-II/I amphipathic clusters in lipoprotein formation, constructs encoding a series of C-terminal apoLp-II/I truncations were created (Figure 2A). C-terminal apoB truncations previously demonstrated that in this related apolipoprotein, the β_1 cluster mediates lipid binding and lipoprotein formation (Yao *et al.*, 1991; Spring *et al.*, 1992; McLeod *et al.*, 1996; Dashti *et al.*, 2002; Shelness *et al.*, 2003; Manchekar *et al.*, 2004). The created apoLp-II/I constructs encode the N-terminal 1290, 1983, 2811 and the complete 3380 aa of *L. migratoria* apoLp-II/I, fused at the C-terminus to a V5/His epitope tag. Excluding the signal sequence, these C-terminal truncations correspond to 38, 59, 84 and 100% of apoLp-II/I, respectively. All constructs contain the complete LLT module, including the full apoLp-II sequence (Figure 2A). ApoLp-II/I-38 encodes the complete LLT module and one third of the β cluster, apoLp-II/I-59 nearly the complete β cluster, whereas apoLp-II/I-84 encompasses all three recognized amphipathic clusters. ApoLp-II/I-100 encodes the complete apoLp-II/I precursor, with all recognized clusters as well as the C-terminal region, including the vWF-D module. ApoLp-II/I-38 expression products have been characterized previously (Chapter 2).

Recombinant expression of the apoLp-II/I-59, apoLp-II/I-84 and apoLp-II/I-100 truncations in Sf9 cells resulted in the secretion of the expected apoLp-II/I cleavage products, as judged from immunoblotting of incubation medium from transfected Sf9 cells (Figure 2B). ApoLp-II was identified by the appearance of immunoreactivity to α-II at a molecular mass of ~72 kDa, for all truncation variants expressed. ApoLp-I was identified by immunoreactivity for α-V5 and α-I, with molecular mass increasing proportional to apoLp-II/I truncation length, *i.e.* ~145, ~220 and ~250 kDa for apoLp-II/I-59, apoLp-II/I-84 and apoLp-II/I-100 truncation variants, respectively. Since the α-II antibody has a mild cross-reactivity to apoLp-I (Jan M. van Doorn, personal communication), a minor immunoreactivity with apoLp-I was observed in some cases. Shared immunoreactivity of α-I, α-II, and α-V5 also identified significant amounts of uncleaved apoLp-II/I in the incubation media for all truncations, with molecular masses of ~230, >250 and >250 kDa, respectively, for the novel constructs. The secretion of uncleaved apoLp-II/I likely relates to the relatively high expression level achieved with the present transfection protocol, as transfection according to a previous protocol did not result in secretion of uncleaved apoLp-II/I-38 protein (Chapter 2). For clarity, as both the cleavage products apoLp-I and apoLp-II as well as their precursor apoLp-II/I are secreted, we will refer to their complement as *apoLp-II/I proteins*.

The relative amount of apoLp-II/I proteins secreted by transfected cells differs between the apoLp-II/I truncation variants. As compared to apoLp-II/I-38 proteins, approximately three times less apoLp-II/I-59 proteins are secreted and about ten times less apoLp-II/I-84 and apoLp-II/I-100 proteins (data not shown). These differences may reflect different expression levels, due to decreased transfection efficiency with larger expression

plasmids, or a decreased secretion rate due to difficulties in the proper folding of apoLp-II/I proteins.

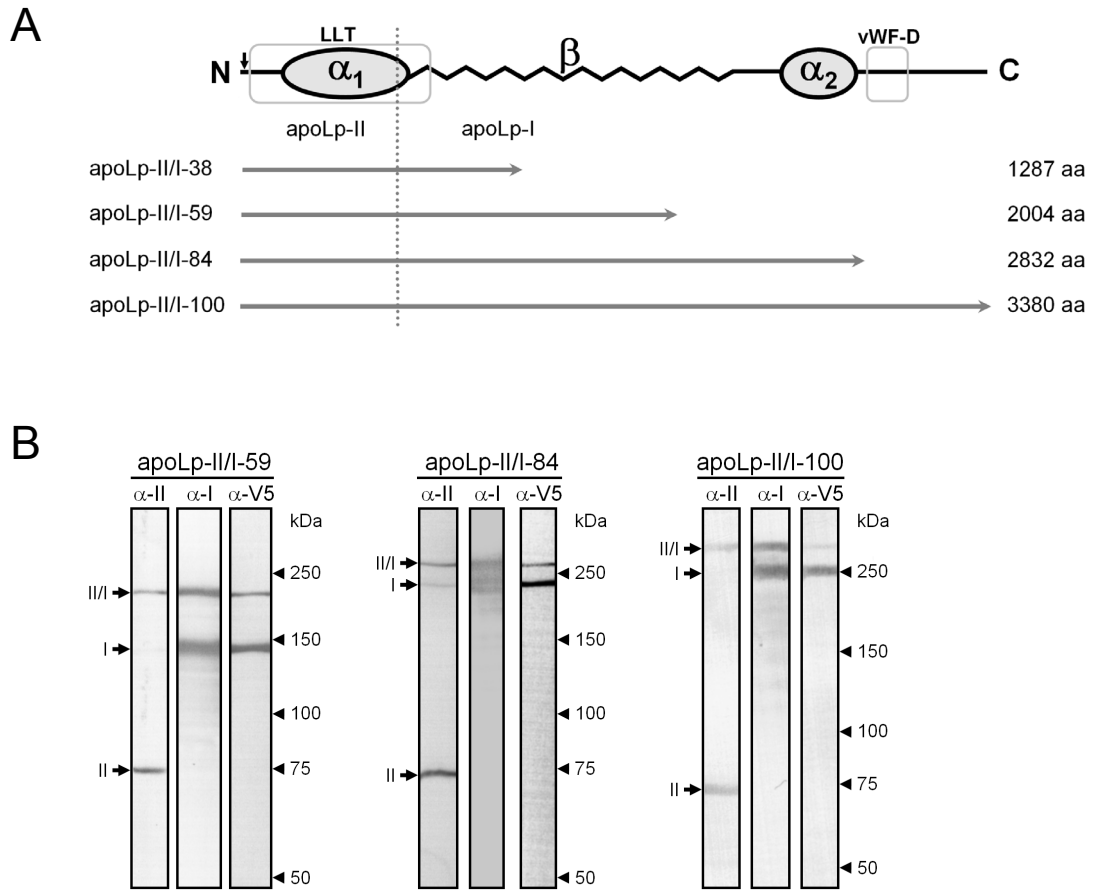


Figure 2. Expression of apoLp-II/I truncations. Distinct expression constructs encoding C-terminal truncation variants of *L. migratoria* apoLp-II/I were created, and subsequently expressed in Sf9 cells. **A**: Comparison of the proteins encoded by the different constructs, indicated by grey horizontal arrows, to the complete *L. migratoria* apoLp-II/I. The names of the constructs at the left of the arrows correspond to the encoded percentage of the complete apoLp-II/I, excluding the signal peptide. The total number of encoded apoLp-II/I amino acids is shown on the right. The recognized structural elements in apoLp-II/I are indicated schematically. Boxes denote the large lipid transfer (LLT) and von Willebrand Factor D (vWF-D) modules, whereas regions enriched in amphipathic α-helices, designated α_1 and α_2, are indicated by ovals, and the region enriched in amphipathic β-strands, named β, by zigzag-lines. The small vertical arrow indicates the predicted signal peptide cleavage site, and the dashed vertical line the location of the site of apoLp-II/I cleavage into apoLp-II and apoLp-I. **B**: Recombinant expression of the novel apoLp-II/I constructs, apoLp-II/I-59, -84 and -100, in Sf9 cells. Immunoblot analysis of incubation medium, using antibodies directed against apoLp-II (α-II) and apoLp-I (α-I and α-V5), demonstrates secretion of (truncated) uncleaved apoLp-II/I as well as its cleavage products apoLp-II and (truncated) apoLp-I (marked by arrows). The position of molecular mass standards is indicated at the right by arrowheads. The relative amount of incubation medium applied for immunoblot was varied according to the apoLp-II/I truncation variant used, to enable detection of variants with a relative low secretion level.

Lipoprotein formation by truncated apoLp-II/I variants

To assess the role of the identified apoLp-II/I amphipathic clusters in lipid binding, we determined the buoyant density of recombinant secreted apoLp-II/I truncation variants using density gradient ultracentrifugation. The apoLp-II/I-100 proteins were all recovered at a buoyant density of 1.18 ± 0.02 g/ml (Figure 4G), suggesting formation of high-density lipoprotein particles. The apoLp-II/I-84 proteins were recovered at a similar buoyant density (1.17 ± 0.02 g/ml; Figure 4E). In contrast, apoLp-II/I-38 proteins were found in the very high-density lipoprotein range only, at 1.24 ± 0.02 g/ml, indicating a very poor lipidation, if any (Figure 4A). ApoLp-II/I-59 proteins floated at an intermediate position of 1.19 ± 0.03 g/ml (Figure 4C). Gel filtration analysis (data not shown) confirmed that apoLp-I and apoLp-II from the apoLp-II/I-59, -84 and –100 variants form lipoprotein complexes as both cleavage products were recovered at the same elution, with estimated molecular weights of 230, 500 and 550 kDa. ApoLp-II/I-38-derived apoLp-I and apoLp-II, however, were recovered at an estimated molecular weight of 100 kDa, below their combined molecular mass, indicating that they do not form a complex. Together, these data demonstrate that the apoLp-II/I-38 truncation cannot form a buoyant lipoprotein, whereas the apoLp-II/I-59, -84, and –100 do. These differences between apoLp-II/I truncation variants in buoyant density are not related to differences in expression level, as apoLp-II/I-38 proteins also failed to produce a high-density lipoprotein at lower expression levels (Chapter 2). Compared to apoLp-II/I-38, apoLp-II/I-59 contains an additional half of the β cluster, amounting to nearly the complete β cluster, whereas apoLp-II/I-84 contains the entire β cluster and the α_2 cluster. Therefore, sequences in the β cluster, and possibly the α_2 cluster, are proposed to enable apoLp-II/I to acquire the lipids required to contitute a high-density lipophorin.

Effect of dMTP on insect lipoprotein formation

The recombinant lipoprotein expression system was subsequently used to investigate a putative role for MTP in insect lipoprotein assembly. A previously reported insect MTP from *D. melanogaster* (dMTP; Sellers *et al.*, 2003), was cloned for co-expression with the apoLp-II/I constructs. As antibodies to dMTP are currently not available, dMTP cDNA was cloned in frame with a V5/His epitope tag to enable detection of the encoded protein. Transfection of this dMTP construct in Sf9 cells resulted in the secretion of dMTP, as evidenced by an α-V5 immunoreactive band at approximately 100 kDa (Figure 3A). This molecular mass is similar to that reported for dMTP upon expression in the mammalian COS-1 cell line (Sellers *et al.*, 2003).

The effect of dMTP co-expression on the secretion of apoLp-II/I proteins by recombinant Sf9 cells was assessed by immunoblot analysis of apoLp-II/I proteins in incubation media. As compared to the control (pIB/V5-His vector co-transfected cells),

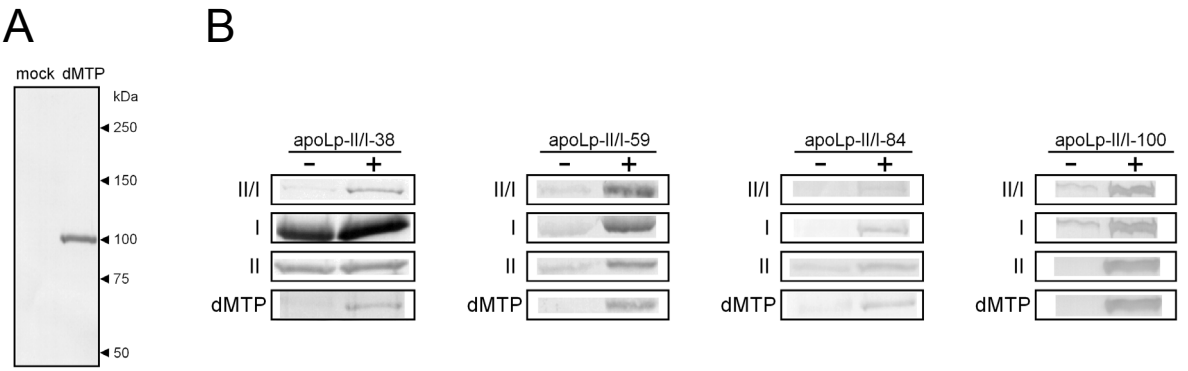


Figure 3. Co-expression of dMTP enhances the secretion of recombinant apoLp-II/I proteins. A: Recombinant expression of dMTP by Sf9 cells. Cells were transiently transfected with control or dMTP construct, and incubation media were assessed for the presence of recombinant dMTP by immunoblotting with α-V5 antibody. **B**: Effect of dMTP co-expression on the secretion of apoLp-II/I-38, apoLp-II/I-59, apoLp-II/I-84, and apoLp-II/I-100 proteins. Sf9 cells were co-transfected with one of the apoLp-II/I constructs as well as dMTP construct (plus sign) or control vector (minus sign). The amount of apoLp-II/I, apoLp-I, apoLp-II and dMTP secreted into the incubation medium was assessed by immunoblotting, using α-II antibody to detect apoLp-II, and α-V5 antibody to detect apoLp-I, apoLp-II/I and dMTP. The relative amount of incubation medium used for immunoblot was varied according to the apoLp-II/I truncation variant used, to enable consistent detection of apoLp-II/I proteins. Results are representative of at least three independent transfections.

co-expression of dMTP with apoLp-II/I-100 in Sf9 cells resulted in a several-fold increase of all secreted apoLp–II/I proteins (Figure 3B). Co-expression of dMTP was verified by α-V5 immunoreactivity at 100 kDa. dMTP impacted the secretion of apoLp-II/I-59 and -84 proteins to a similar extent, whereas the increase in secretion of apoLp-II/I-38 proteins was clearly lower (Figure 3B). Thus, an insect MTP enhances the secretion of recombinant apoLp-II/I proteins by Sf9 cells.

To assess the effect of MTP on the lipidation of insect lipoproteins, we determined the effect of dMTP co-expression on the buoyant density of the secreted apoLp-II/I proteins, *i.e.* apoLp-I, apoLp-II and the uncleaved precursor apoLp-II/I. Co-expression of dMTP was confirmed by α-V5 immunoreactivity at 100 kDa (only shown in Figure 4H). dMTP co-expression clearly affected the buoyant density of the apoLp-II/I proteins for all truncation variants, except apoLp-II/I-38 (Figure 4A, B), since buoyant density decreased from 1.19 ± 0.03 g/ml to 1.15 ± 0.01 g/ml for apoLp-II/I-59 proteins (Figure 4C, D), from 1.17 ± 0.02 g/ml to 1.14 ± 0.01 g/ml for apoLp-II/I-84 proteins (Figure 4E, F) and from 1.18 ± 0.02 g/ml to 1.14 ± 0.02 g/ml for apoLp-II/I-100 proteins (Figure 4G, H). For all truncation variants, the uncleaved apoLp-II/I displayed the same buoyant density distribution as its cleavage products apoLp-I and apoLp-II, confirming that apoLp-II/I cleavage is not an indispensable step in high-density lipophorin biogenesis (Chapter 2).

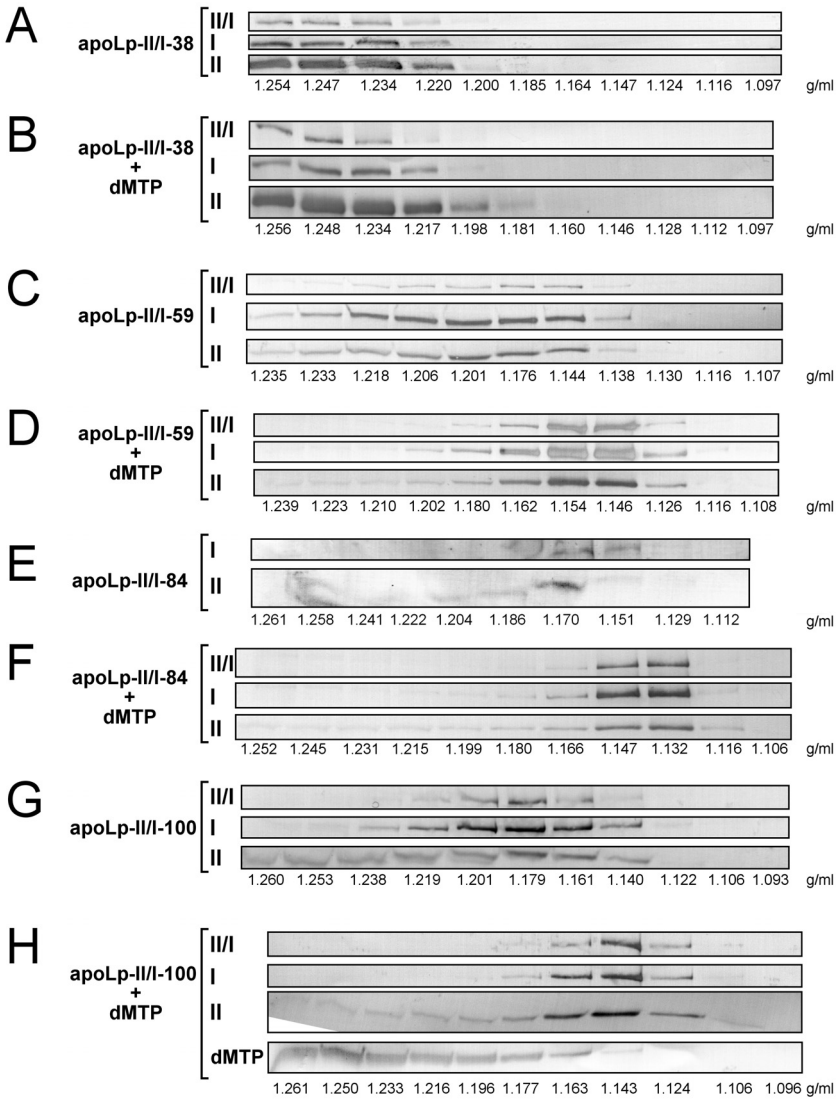


Figure 4. Co-expression of dMTP selectively decreases the buoyant density of recombinant apoLp-II/I proteins. Sf9 cells were transiently transfected with one of the different apoLp-II/I constructs, along with dMTP or control vector, and incubation medium was submitted to density gradient ultracentrifugation. The resulting gradients were divided in fractions that were analyzed for apoLp-II/I proteins by immunoblotting with α-V5 and α-II antibodies. The relative amount of fractions used for immunoblot was varied between the gradients according to secretion levels, to enable consistent detection of apoLp-II/I proteins. Panels show immunoblotting results on the distribution of apoLp-I, apoLp-II and apoLp-II/I among the fractions of each gradient, and are accordingly marked by I, II and II/I, respectively. For all gradients, the buoyant density of each fraction is indicated under the lower panel. Subfigures A to H show the density distribution of apoLp-II/I proteins for different sets of transfected constructs, representative of at least three independent transfections. **A**: apoLp-II/I-38 and control; **B**: apoLp-II/I-38 and dMTP; **C**: apoLp-II/I-59 and control; **D**: apoLp-II/I-59 and dMTP; **E**: apoLp-II/I-84 and control; **F**: apoLp-II/I-84 and dMTP; **G**: apoLp-II/I-100 and control; **H**: apoLp-II/I-100 and dMTP. In H, a panel has been included that shows the distribution of dMTP in the density gradient, representative for its co-expression with any of the apoLp-II/I truncation variants.

dMTP is recovered in the higher density range of the gradient, mostly at a density above 1.20 g/ml (Figure 4H), demonstrating that it is not or only poorly lipidated, and that it is not secreted by virtue of association with apoLp-II/I lipoproteins, as also indicated by its expression by Sf9 cells in the absence of apoLp-II/I co-transfection (Figure 3A). In summary, the present findings show that, in a recombinant expression system, an insect MTP promotes the secretion as well as the lipidation of apoLp-II/I proteins.

Discussion

It is well established that the two apolipoproteins of insect lipoprotein, apoLp-I and apoLp-II, are derived from cleavage of their precursor protein apoLp-II/I (Weers *et al.*, 1992; Chapter 2). Yet the mechanism that allows apoLp-II/I proteins to acquire lipids and accordingly form a lipoprotein remains to be characterized. In the present study, we identified clusters enriched in either amphipathic α-helices or amphipathic β-strands in apoLp-II/I, and demonstrated the requirement of the β cluster for biogenesis of a buoyant lipoprotein, using a novel, insect cell-based recombinant expression system for insect lipoprotein. ApoLp-II/I may acquire at least part of its bound lipids from MTP, as we found *D. melanogaster* MTP to promote the secretion of apoLp-II/I proteins and to enhance their lipidation to a high-density lipoprotein. These effects of dMTP depend on the presence in apoLp-II/I of the larger part of the predicted amphipathic β cluster. Thus, insect lipoprotein biogenesis appears to involve the same major structural elements as mammalian apoB lipoprotein biogenesis. Moreover, in accordance with our earlier study (Chapter 2), we find that cleavage of apoLp-II/I is not a prerequisite for lipophorin biogenesis, since uncleaved apoLp-II/I was also lipidated and displayed a buoyant density similar to native HDLp.

LOCATE analysis predicts insect apoLp-II/I to contain three regions enriched in either amphipathic α-helices or amphipathic β-strands, that are organized along the polypeptide as N-α_1-β-α_2-C. This tripartite organization is reminiscent of a truncated form of its vertebrate homolog apoB, that was predicted to contain a pentapartite organization, viz. N-α_1-β_1-α_2-β_2-α_3-C (Segrest *et al.*, 1994, 1998). However, the β cluster of apoLp-II/I is significantly longer than either of the two β clusters in apoB, whereas apoLp-II/I's α_2 cluster is only about two third the size of the α_2 or α_3 clusters in apoB. Based on their sequence similarity in the N-terminal LLT module, which includes the α_1 cluster and a small part of the β_1 cluster, apoB and apoLp-II/I have previously been recognized as homologs (Babin *et al.*, 1999; Mann *et al.*, 1999). Therefore, the amphipathic clusters C-terminal from the LLT module of apoB and apoLp-II/I may also share a common evolutionary origin. From this perspective, one speculative scenario to explain the present differences between apoB and apoLp-II/I amphipathic clusters would be that the β_2 and α_3

clusters in apoB have arisen from a duplication of the β_1 and α_2 clusters. On the other hand, the amphipathic structures outside of the LLT module in apoB and apoLp-II/I may have evolved independently, and thus represent an example of convergent evolution. The lack of clear sequence conservation between apoB and apoLp-II/I C-terminal to the conserved N-terminal LLT module does not allow for a conclusive statement on the origin of the amphipathic cluster organization at present.

Recombinant expression of apoLp-II/I proteins demonstrates the importance of the β cluster region in the biogenesis of insect lipoprotein, as only C-terminal truncations of apoLp-II/I with at least the larger part of the β cluster were secreted as buoyant lipoproteins. In addition, the α_2 cluster may also enhance lipid acquisition, as suggested by the slightly lower density of apoLp-II/I-84 lipoprotein, compared to apoLp-II/I-59 lipoprotein in the presence as well as absence of dMTP co-expression. On the other hand, apoLp-II/I-84 also contains an additional ~250 aa of the β cluster region that might mediate the additional lipid association as well. Inclusion of the region C-terminal to the α_2 cluster hardly affected lipoprotein density. The vWF-D module in this region apparently does not function in lipid binding. Thus, the β cluster region in apoLp-II/I is essential to lipoprotein formation. The acquired lipids are likely bound and stabilized by the β-strands of high lipid-binding affinity predicted in this region. We propose that the apoLp-II/I region N-terminal to the β cluster, *i.e.* most of the LLT module, facilitates lipid acquisition by the β cluster, similar to the homologous region in apoB, which interacts with MTP and constitutes part of the lipid-binding cavity responsible for initial lipid acquisition (Gretch *et al.*, 1996; Mann *et al.*, 1999; Shelness *et al.*, 1999; Manchekar *et al.*, 2004).

Similar to the β cluster region in apoLp-II/I, the β_1 cluster in apoB is involved in lipoprotein formation. Inclusion of increasingly more of the β_1 cluster results in lipoproteins of decreased buoyant density that accordingly contain more lipid (Yao *et al.*, 1991; Spring *et al.*, 1992; Dashti *et al.*, 2002; Shelness *et al.*, 2003; Manchekar *et al.*, 2004). However, unlike apoLp-II/I in the present study, apoB truncations that include hardly any β_1 cluster sequence already formed high-density lipoproteins. Moreover, C-terminal truncations of apoB including the complete β_1 cluster can form lipoproteins in the very low-density range (d<1.006 g/ml) *in vivo* (for review see Davidson and Shelness, 2000), binding far more lipids than the apoLp-II/I high-density lipoproteins formed in insects (Venkatesh *et al.*, 1987; Capurro and De Bianchi, 1990; Weers *et al.*, 1992; Van Heusden *et al.*, 1998). These differences in lipid acquisition may reflect a limited lipid binding capacity of apoLp-II/I's β cluster as compared to apoB's β_1 cluster, but might also result from differences in the mechanism of apoLp-II/I and apoB lipoprotein assembly, the cell lines used (Graham *et al.*, 1991; Gretch *et al.*, 1996; Manchekar *et al.*, 2004), or incubation conditions, *e.g.* supplementation with exogenous lipids.

The amphipathic α and β clusters of apoB are believed to play different roles in lipid association. The β clusters are predicted to form nearly continuous amphipathic β-sheets of sufficiently high lipid affinity to nearly irreversibly associate with lipids, whereas the amphipathic α-helices of the α_2 and α_3 clusters may reversibly associate or dissociate from the lipid core upon changes in lipid content (Segrest *et al.*, 2001; Wang *et al.*, 2003; Wang and Small, 2004). Amphipathic β-strands and α-helices may play similar roles in the dynamics of insect lipoprotein. Strikingly, lipid transfer can exceed the lipid-binding capacitity of HDLp during extreme lipid transport conditions in the insect, *e.g.* during sustained flight. Several molecules of a specialized exchangeable apolipoprotein consisting of five reversibly lipid-binding amphipathic α-helices, apoLp-III, then bind and stabilize lipophorin (for review see Ryan and Van der Horst, 2000; Van der Horst *et al.*, 2001). In view of the present results, apoLp-III may be regarded as a reservoir of lipid-binding amphipathic α-helices that compensates for the relative small amount of amphipathic α-helices present in apoLp-I and -II as compared to apoB.

Following cleavage of apoLp-II/I into apoLp-I and apoLp-II, the β cluster is almost entirely situated in apoLp-I. From this perspective, the finding that inclusion of the β cluster enables HDLp biogenesis suggests that apoLp-I, and not apoLp-II, binds the vast majority of lipids. This finding is in accordance with HDLp dissociation experiments in which over 98% of the total lipid in lipophorin remained associated to apoLp-I (Kawooya *et al.*, 1989). Nonetheless, apoLp-I and apoLp-II most probably cannot be regarded as independent entities during insect lipoprotein biogenesis, as lipidation likely occurs prior to cleavage of apoLp-II/I, and apoLp-I and apoLp-II regions are intimately linked within the structure of the LLT module (Chapter 2).

Whereas insect lipoprotein appears to bind lipids via the amphipathic β cluster in apoLp-II/I, our results also suggest that insect MTP enhances the lipid transfer to apoLp-II/I. Co-expression of dMTP clearly stimulated the lipidation of apoLp-II/I-59, apoLp-II/I-84 and apoLp-II/I-100, as judged from their shift to a lower buoyant density. This stimulation is dependent on the presence of the larger part of the β cluster, as apoLp-II/I-38 failed to form a buoyant lipoprotein regardless of dMTP co-expression. In addition, dMTP co-expression resulted in a several-fold increase in the secretion of recombinant apoLp-II/I proteins, for all truncation variants studied. Similarly, MTP co-expression enhances the secretion of apoB truncations longer than the N-terminal 1000 aa of apoB. Unlike our present findings for apoLp-II/I, mammalian cells secrete hardly any apoB in the absence of MTP expression (Leiper *et al.*, 1994; Wang *et al.*, 1997; Dashti *et al.*, 2002; Shelness *et al*., 2003; Manchekar *et al.*, 2004). The secretion and limited lipidation of apoLp-II/I truncations by Sf9 cells in the absence of dMTP co-expression may indicate that MTP is not an absolute requirement in insect HDLp biogenesis. However, it cannot be excluded that Sf9 cells express an endogenous MTP, especially since Sf9-derived cells have previously been reported to possess limited apoB lipoprotein assembly capacity

(Gretch *et al.*, 1996). In any case, dMTP co-expression clearly enhanced the cellular capacity of Sf9 cells to produce HDLp, suggesting that MTP functions in insect lipoprotein biogenesis *in vivo* as well.

In mammals, MTP promotes the secretion of apoB lipoproteins by decreasing apoB misfolding, intracellular retention and degradation. ApoB is particularly prone to degradation during the co-translational step of lipidation, and MTP associates with apoB's LLT module to promote the lipidation and folding of apoB's amphipathic sequence regions at this step (for reviews see Davidson and Shelness, 2000; Shelness and Sellers, 2001; Mahmood Hussain *et al.*, 2001, 2003). It remains to be determined whether MTP directly promotes either proper folding or net lipidation of nascent apoB (for review see Mahmood Hussain *et al.*, 2003), and the ability of MTP to transfer lipids may only be relevant for formation of intralumenal neutral lipid droplets during the second post-translational lipidation step. The concomitant increased secretion and decreased buoyant density of recombinant apoLp-II/I upon dMTP co-expression suggests that dMTP also promotes insect lipoprotein secretion by enhancing apoLp-II/I lipidation, possibly by preventing intracellular degradation and/or retention of nascent apoLp-II/I lipoproteins. However, the mechanism by which dMTP promotes insect lipoprotein biogenesis remains to be established.

Given the specific features of insect lipoproteins and lipid transport, it appears unlikely that the mechanism of apoLp-II/I lipoprotein biogenesis is identical to apoB lipoprotein biogenesis. Importantly, upon secretion, insect lipoprotein contains far less lipid molecules than apoB lipoproteins, *i.e.* several hundred *vs.* over a thousand, respectively. ApoB acquires the majority of these lipids after its co-translational lipidation by fusion of the primordial lipoprotein with an intralumenal neutral lipid droplet, resulting in a very low-density lipoprotein (for reviews see Shelness and Sellers, 2001; Mahmood Hussain *et al.*, 2001; Borén and Olofsson, 2005). Therefore, the relatively modest lipidation of apoLp-II/I to a high-density lipoprotein could result from the absence of this second step in insect lipoprotein biosynthesis. In addition, the major lipid components in newly biosynthesized insect HDLp and mammalian apoB lipoprotein are DAG and TAG, respectively (Weers *et al.*, 1992; for review see Ryan and Van der Horst, 2000), suggesting that the lipoprotein producing cells of insects and mammals differ in their intracellular lipid mobilization. The lipid transfer activity of MTP may have been optimized accordingly, as suggested by the far lower *in vitro* TAG transfer activity of dMTP as compared to human MTP (Sellers *et al.*, 2003). Furthermore, we observed dMTP to be secreted efficiently by insect cells. In contrast, mammalian MTP is not secreted but retained in the secretory pathway, due to its stable interaction with the ER-resident chaperone protein disulfide isomerase (PDI) (Wetterau *et al.*, 1990). Therefore, the observed secretion of dMTP may reflect the lack of interaction with PDI, either by an insufficient amount of PDI or by an inability of dMTP to heterodimerize with PDI. Moreover, the secretion of recombinant dMTP is of special interest since insect

hemolymph contains large amounts of the lipid transfer particle LTP (for reviews see Ryan and Van der Horst, 2000; Van der Horst and Ryan, 2004). LTP enables the lipoprotein-mediated lipid transport within insects as it mediates the vectorial redistribution of lipids among lipoproteins and between lipoproteins and tissues. LTP is a very high-density lipoprotein that contains three distinct apoproteins, and is produced and secreted by the fat body. Based on the similar activity and the secretion of dMTP by recombinant insect cells, we speculate that MTP is a component of the LTP complex.

The involvement of MTP in lipoprotein assembly in insects as well as mammals implies a common evolutionary origin for lipoprotein formation, prior to the divergence of deuterostome and protostome lineages of animals. This prompts the question when and how this system arose in evolution. Strikingly, the lipid transfer protein MTP as well as its substrates apoLp-II/I and apoB are members of the homologous family of large lipid transfer proteins (LLTP), identified by the presence of an LLT module (Babin *et al.*, 1999; Mann *et al.*, 1999). Another LLTP is the yolk protein vitellogenin (Vtg) (Babin *et al.*, 1999). Vtg was previously assumed to be ancestral to the other LLTP, based on its nearly ubiquitous presence among animal classes (mammals excepted) and limited lipid binding. However, the biosynthesis of vertebrate and invertebrate Vtg was recently shown to be MTP-dependent (Shibata *et al.*, 2003; Sellers *et al.*, 2005), suggesting that MTP is the ancestral member of the LLTP gene family and may function in the biogenesis of all of its evolutionary descendants. Our findings concur with latter hypothesis, as we find that in addition to apoB and vitellogenin, MTP also stimulates the biosynthesis of apoLp-II/I.

In conclusion, we propose that lipoprotein biogenesis in insects depends on the same major structural elements as in mammals, *i.e.* MTP as well as amphipathic structures in the carrier apolipoprotein. This implies an early metazoan origin for the current mechanisms of lipoprotein assembly. Many questions on the origin of LLTP-derived lipoproteins remain to be answered, including whether MTP is ancestral to the other LLTP, whether apoLp-II/I is the insect apoB ortholog, and on the origin of amphipathic clusters in apoB and apoLp-II/I. The availability of multiple complete vertebrate and invertebrate genome sequences may provide further insight into the diversity and molecular evolution of LLTP, and therefore the evolution of lipoproteins. The early metazoan origin of lipoprotein assembly implies ample time for divergence and adaptations, and specific features of insect lipoprotein, *e.g.* binding less and different lipids, may be understood from this perspective. Thus, further research on the mechanism of insect lipoprotein biosynthesis may provide novel venues to manipulate the biosynthesis of apoB lipoproteins, and accordingly to the treatment of atherosclerosis and other lipid disorders.

Acknowledgements

The authors thank Jan M. Van Doorn and Marcelle A.M. Kasperaitis for technical assistance, and Jere P. Segrest and Martin K. Jones (Department of Medicine, University of Alabama at Birmingham Medical Center) for providing the program LOCATE. We thank Gregory S. Shelness (Department of Pathology, Wake Forest University School of Medicine) for valuable discussions, and Carol C. Shoulders (MRC Clinical Sciences Centre, Imperial College London) for critical reading of the manuscript.

Chapter 4

Molecular diversity and evolution of the large lipid transfer protein family

Marcel M.W. Smolenaars[1], Ole Madsen[2], Kees W. Rodenburg[1], and Dick J. van der Horst[1]

[1] Biochemical Physiology, Department of Biology and Institute of Biomembranes, Utrecht University, The Netherlands

[2] Department of Biochemistry, Faculty of Science, Radboud University Nijmegen, The Netherlands

to be submitted

Abstract

The family of large lipid transfer proteins (LLTP) consists of homologous lipid-binding proteins of high physiological and pathological importance, including vitellogenin (Vtg), apoB, apoLp-II/I, and MTP. To attain a view on the diversity of LLTP, available sequence databases were searched for proteins that contain the family-defining large lipid transfer (LLT) module. LLTP could only be identified in metazoan phyla, with represented LLTP subsets varied among taxa. Two novel LLTP were recognized in insects, melanin-engaging protein (MEP) as well as an apoLp-II/I-related protein (ARP). Phylogenetic analysis of conserved sequence blocks in the LLT module revealed three major subfamilies of LLTP: (1) vertebrate and invertebrate MTP, (2) vertebrate apoB, insect apoLp-II/I, insect ARP, mite M-177 allergen, and Vtg from decapodan Crustacea, (3) Vtg (excluding decapodan Vtg), decapodan clotting protein and insect MEP. Strikingly, the Vtg from decapodan Crustacea is not Vtg-like, but an apoB-like LLTP, exemplifying the functional flexibility of LLTP. The evolution and diversification of LLTP is discussed in relation to the apparent variation in LLTP functionality.

Introduction

Lipid transport potentially poses a problem to multicellular organisms, as lipids are hydrophobic and aggregrate in an aqueous environment. In animals, members of the family of large lipid transfer proteins (LLTP) enable lipid transport, as these proteins can bind and bundle up to hundreds of lipid molecules (for reviews see Hevonoja *et al.*, 2000; Segrest *et al.*, 2001). In this way, LLTP constitute lipoproteins that endow the uptake and transport of a multitude of lipid species. Accordingly, LLTP have been implicated in animal development (Panakova *et al.*, 2005), vitellogenesis (for reviews see Sappington and Raikhel, 1998a; Walzem *et al.*, 1999; Romano *et al.*, 2004), immunity (Hall *et al.*, 1999; Yeh *et al.*, 1999; Brozovic *et al.*, 2004; Dougan *et al.*, 2005), as well as ageing and lifespan regulation (Brandt *et al.*, 2005; Amdam *et al.*, 2005). Moreover, lipoproteins constituted by LLTP are a major causitive agent of atherosclerosis in man, and involved in other metabolic disorders such as obesity and diabetes mellitus 2 (for reviews see Olofsson and Borén, 2005; Shoulders and Shelness, 2005; Taskinen, 2005).

LLTP have been identified in most metazoan phyla, and were recognized to include apolipoprotein B (apoB), apolipophorin-II/I (apoLp-II/I), vitellogenin (Vtg), and microsomal triglyceride transfer protein (MTP) (Babin *et al.*, 1999). All LLTP bind lipids, albeit in different amounts and compositions, and for different purposes. ApoB is present in vertebrates and mediates post-prandial lipid uptake as well as lipid redistribution in the body (for review see Frayn, 2003), as hundreds of lipid molecules may be grouped and stabilized by a single apoB (for reviews see Hevonoja *et al.*, 2000; Segrest *et al.*, 2001).

Like apoB, the insect apoLp-II/I functions as a general-purpose lipid transporter. However, in contrast to apoB, apoLp-II/I additionally mediates the transport of stored lipid for energy generation during muscular activity (for reviews see Ryan and Van der Horst, 2000; Van der Horst and Ryan, 2004). Moreover, the lipoprotein constitutes a major coagulogen in insect hemolymph (see Chapter 2). Vtg is the major yolk protein found in females of most egg-laying animals, non-mammalian vertebrates as well as invertebrates, supplying the developing oocyte with nutrients, including lipids (for reviews see Sappington and Raikhel, 1998a; Romano *et al.*, 2004). MTP is present in vertebrates and invertebrates and has been shown to enhance the biosynthesis of other LLTP, as demonstrated for apoB, apoLp-II/I and Vtg (Shibata *et al.*, 2003; Shelness *et al.*, 2005; Shelness and Ledford, 2005; Chapter 3). MTP and most Vtg can bind relative little lipids, as compared to apoB and apoLp-II/I. Vertebrate Vtg are known to form homodimers (Anderson *et al.*, 1998; Kollman and Quispe, 2005). During their biogenesis, apoLp-II/I and several invertebrate Vtg are cleaved, yet the resulting polypeptides remain in complex, as illustrated by the presence of the apoLp-II/I cleavage products apoLp-I and –II in insect lipoprotein (Weers *et al.*, 1993; Sappington and Raikhel, 1998a; Chapters 2 and 3).

LLTP are defined by their large lipid transfer (LLT) module, comprising the N-terminal ~900 amino acid residues, that has been conserved in sequence and structure among LLTP (Raag *et al.*, 1988; Anderson *et al.*, 1998; Babin *et al.*, 1999; Mann *et al.*, 1999; Segrest *et al.*, 1999; Read *et al.*, 2000; Thompson and Banaszak, 2002; Richardson *et al.*, 2005; Chapter 2). The more conserved N-terminal 600 residues of the LLT module form a barrel-like β-sheet (βC) as well as a coiled horsehoe-shaped α-helical bundle, with α-helices arranged in two layers. The less conserved C-terminal portions encode two amphipathic β sheets (βA and βB) that constitute the larger part of the lipid-binding cavity. Studies on apoB lipoprotein biosynthesis indicate that the LLT module provides the structural basis for initial acquisition and binding of lipids (for review see Segrest *et al.*, 2001; Shelness and Sellers, 2001). Several modifications within the LLT module have been observed. ApoLp-II/I is post-translationally cleaved within the LLT module (Chapter 2), as are some insect Vtg (for review see Sappington and Raikhel, 1998a). Several insect Vtg also contain repeats of serine residues (*i.e.* polyserine tracts) within the LLT module, that may be phosphorylated and accordingly bind ions (for review see Sappington and Raikhel, 1998a).

Apart from their common LLT module, LLTP differ to a great extent. MTP is the smallest LLTP and contains little more sequence than the LLT module. In contrast, vertebrate apoB spans ~4500 amino acids (aa), insect apoLp-II/I ~3350 aa, and Vtg ~1700 aa (see Shelness and Ledford, 2005). In addition to the LLT module, Vtg and apoLp-II/I have been demonstrated to contain a single von Willebrand Factor D (vWF-D) module near their C-terminus, unlike MTP and apoB (Babin *et al.*, 1999). The function of this domain remains elusive. Apart from the LLT and vWF-D modules, hardly any sequence

similarity can be observed among apoB, apoLp-II/I, vitellogenin and MTP (Babin *et al.*, 1999). Structural predictions, however, indicate that apoB and apoLp-II/I contain clusters enriched in either amphipathic α-helices or amphipathic β-strands, organized along these proteins as N-α_1-β_1-α_2-β_2-α_3-C and N-α_1-β-α_2-C, respectively (Segrest *et al.*, 1994, 1998; Chapter 3). The N-terminal amphipathic α-helix cluster corresponds to the α-helical bundle of the LLT module, whereas part of the adjacent amphipathic β-strand cluster constitutes the lipid-binding cavity of the LLT module. The β cluster appears to harbor most of the lipid-binding capacity in apoB (for review see Segrest *et al.*, 2001) as well as apoLp-II/I (Chapter 3).

LLTP were previously demonstrated to be derived from a common ancestor, based on similarity and ancestral exon boundaries (Babin *et al.*, 1999). However, the evolutionary relationships among LLTP and the nature of their common ancestor could not be established clearly, due to a lack of sequence information (Babin *et al.*, 1999). Vtg has been assumed to be ancestral to the other LLTP, based on its ubiquitous presence among animal phyla and important role as a yolk protein in reproduction. However, the recent demonstration that MTP facilitates apoB, apoLp-II/I and Vtg biosynthesis in vertebrates as well as invertebrates, suggests that MTP may be the ancestral family member of LLTP (Shibata et al., 2003; Shelness et al., 2005; Shelness and Ledford, 2005; Chapter 3). The currently available LLTP sequences, from molecular cloning studies as well as eukaryote genome sequences, may provide further insight into the molecular diversity as well as evolution of LLTP and their specific elements, *i.e.* the LLT and vWF-D modules as well as amphipathic clusters. Therefore, we presently explore the molecular diversity, in sequence and structure, as well as the phylogeny of the LLTP family.

Methods

Database searches and LLTP selection. The non-redundant GenBank protein database was searched for LLTP by PSI-BLAST and BLASTP (version 2.2.11; Altschul *et al.*, 1997) with the BLOSUM 62 matrix (Henikoff and Henikoff, 1993), using different regions from previously identified LLTP, selected on type and phylogenetic spread. Results from multiple and iterative searches were compiled. Candidate hits were identified as LLTP if they shared significant similarity with LLT modules from previously identified LLTP, *i.e.* they were recovered at a significant score ($E<0.001$) upon BLASTP searching with at least two types of LLTP (Vtg, apoLp-II/I, apoB, or MTP), and reciprocal BLASTP searching recovered the LLTP they had previously been detected with. Hits resulting from strong bias in amino acid composition, for example due to polyserine tracts, were removed. Borderline cases were further investigated by literature analysis. Identified LLTP were compiled and selected for further analysis, based on type, taxonomic

distribution, unique properties and novelty. Whenever availiable, in each phylum two to three sequences of each LLTP type (established in searches of whole genomes) were selected, provided their overall similarity was less than 90%.

Analysis of modular organization and structural elements. Conserved protein domains in selected LLTP were identified using the Conserved Domain Database and Search Service with Reverse Position Specific BLAST (version 2.04; Marchler-Bauer *et al.*, 2005) at NCBI. Polyserine and polyglutamine tracts were identified using WinPep (Hennig, 1999). Literature was analyzed to assess LLTP for the probable and established presence of post-translational cleavage sites. The distribution of amphipathic α-helices and β-strands in LLTP was analyzed using LOCATE (Segrest *et al.*, 1994, 1998, 1999, 2001), with settings as described previously (Chapter 3). The crystal structure of lipovitellin (the N-terminal cleavage product of Vtg) from lamprey (PDB accession 1LSH; Thompson and Banaszak, 2002) was visualized with DeepView (version 3.7; Guex and Peitsch, 1997) to illustrate the conserved and variable sequence regions in the LLT module.

Multiple alignment and phylogenetic analysis. The LLT modules of selected LLTP were manually aligned using GeneDoc (version 2.6.002) by sequential alignment of individual sequences, using the alignments proposed by individual BLAST searches. The alignment was subsequently optimized by reference to previously published alignments (Spieth *et al.*, 1991; Chen *et al.*, 1997; Babin *et al.*, 1999; Mann *et al.*, 1999; Read *et al.*, 2000) as well as alignments of relatively similar LLTP created by T-Coffee (http://igs-server.cnrs-mrs.fr/Tcoffee; Notredame *et al.*, 2000; Poirot *et al.*, 2003). The allowed conservative substitutions were defined as follows: A, G; S, T; E, D; R, K, H; Q, N; V, I, L, M; Y, F; W; P; C. For phylogenetic analysis, only the identified conserved sequence blocks within the LLT module were used. The best fitting model for amino acid sequence evolution was selected using ProtTest (Abascal *et al.*, 2005), and found to be WAG + I + Γ_4 for the present LLT module alignment. This model was used in subsequent phylogenetic analyses. Maximum likelihood (ML) analysis was performed using PHYML 2.4.4 (http://atgc.lirmm.fr/phyml; Guindon *et al.*, 2005), with a BIONJ starting tree and optimalization of gamma shape parameters and proportion of invariable sites during analyses. Node stability was estimated by 100 replicates of non-parametric bootstrapping. Bayesian phylogenetic analysis was performed with MRBAYES (version 3.1.1; Huelsenbeck and Ronquist, 2001). Four Markov chains were run simultaneously for 1,000,000 generations, with initial equal probabilities for all trees and starting with a random tree. Tree sampling was done each 100 generations and the consensus tree with posterior probabilities was calculated after removal of the first 500 trees ("burn in" as determined from the likelihood values). Bayesian analyses was run twice to check consistency of results, and resulted in the same topologies, except for the local placement

of the mite M-177 and zebrafish Vtg, with a maximum observed difference in posterior probability of 2%. Posterior probabilities are indicated as percentages in this study.

Results

Identification of LLTP family members

Multiple and iterative searches identified a total of 234 unique LLTP, including all previously acknowledged LLTP types (Babin *et al.*, 1999), *i.e.* apoLp-II/I, apoB, Vtg, and MTP. In addition, several differently named proteins found in arthropods, were recognized as LLTP, *i.e.* the clotting protein (CP) from decapodan Crustacea (*e.g.* shrimp, crayfish, crab; Hall *et al.*, 1999; Yeh *et al.*, 1999), melanin-engaging protein (MEP) from the insect *Tenebrio molitor* (Lee *et al.*, 2000), and M-177 from dust mite (Arachnida; Epton *et al.*, 1999). Moreover, several Vtg of special evolutionary interest were identified, *i.e.* from oyster (Mollusca; Matsumoto *et al.*, 2003), coral (Cnidaria; accession BAD74020.1, Hayakawa *et al.*, unpublished), water flea (Kato *et al.*, 2004), and decapodan Crustacea (*e.g.* Abdu *et al.*, 2003; Tsang *et al.*, 2003). Despite their failure to meet the present identification criterium of being recovered upon BLAST search with multiple LLTP types, the putative MTP sequences for *D. melanogaster* and *C. elegans* were assigned as LLTP, because of their functional and sequence similarity to vertebrate MTP (Chapter 3; Sellers *et al.*, 2003; Shibata *et al.*, 2003). Unfortunately, LLTP remain to be identified in most invertebrate phyla, such as the polyphyletic group of Porifera. Although several hits on bacteria, plants and unicellular eukaryotes were obtained, none of these met the identification criteria. Thus, no LLTP could be identified in non-metazoan species.

The availability of annotated genomes allowed for the analysis of the complete set of LLTP family members present in a single species. We limit the analysis to the LLTP found in the *Homo sapiens*, *Danio rerio*, *Drosophila melanogaster* and *Caenorhabditis elegans* genomes as the assembly and annotation of these genomes has reached a matured level. More recent animal genome and cDNA projects are used for reference only.

Our searches identified four distinct LLTP in the fruitfly *D. melanogaster*, including the previously demonstrated apoLp-II/I and MTP. Unlike other insects, we could not identify Vtg in the fruitfly, in accordance with the finding that fruitfly and other higher Diptera rely on another unrelated protein as major yolk constituent (for review see Bownes, 1992; Sappington, 2002). In addition to apoLp-II/I and MTP, we identified a putative LLTP (acc. AAM50129) with by far the highest similarity to *T. molitor* MEP (acc. BAB03250; Lee *et al.*, 2000). This apparent MEP from *D. melanogaster* is supported by two full-length cDNA sequences (acc. NM_142244.1 and NM_169676.1). In addition to *D. melanogaster* (order Diptera) and *T. molitor* (Coleoptera), MEP

homologs were also identified in the genome of the honey bee *Apis mellifera* (Hymenoptera; acc. XP_395423.2), suggesting that MEP may be present in most insects. In addition to MEP, another novel LLTP was identified in the *D. melanogaster* genome that had highest similarity to apoLp-II/I and accordingly is named apoLp-II/I-related protein (ARP) in the present study. This LLTP is supported by EST sequences (acc. AA140964, AA440024, AI260281, AW942336, CO155780, CO337900, CO275893, CO330605, CO300300, CO306852, AI944515) as well as two cDNA sequences that reflect two splice variants, differing by the presence (acc. AAS64667.1) or absence (acc. AAF52796.1) of 29 additional residues after residue 451. In addition to *D. melanogaster*, an as yet prematurely annotated ARP homolog was predicted in the genome of *A. mellifera* (Hymenoptera; acc. XP_394269.2), indicating that ARP may be present in most insects. Thus, the fruitfly *D. melanogaster* has at least 4 distinct LLTP genes, *i.e.* MTP, apoLp-II/I, MEP, and ARP. Unlike this higher Dipteran, most insects additionally have one to three relative closely related Vtg.

Analysis of the annotated *H. sapiens*, *D. rerio*, and *C. elegans* genomes did not yield previously unacknowledged LLTP. Apparently man has only two LLTP, apoB and MTP. Similarly, apoB and MTP were the only LLTP identified in the current annotations of *Bos taurus*, *Canis familiaris*, *Mus musculus* and *Rattus norvegicus* genomes, supporting the view that apoB and MTP are the only LLTP in mammals. In addition to apoB and MTP, non-mammalian vertebrates were found to possess several relatively closely related Vtg. Interestingly, two apoB protein sequences of 3730 and 4418 aa were predicted to be encoded by two distinct genes in the zebrafish (acc. CAH68927 and XP_694827, respectively). Literature survey indicates that it remains to be established whether both apoB variants are present in zebrafish lipoproteins. The nematode *C. elegans* has two distinct LLTP types, a MTP (Shibata *et al.*, 2003) as well as six relatively closely related Vtg genes. Thus, organisms from distinct taxa contain a specific subset of LLTP members.

Modular organization of LLTP

Based on their taxonomic distribution, diversity and novelty, a subset of the identified LLTP was selected for further structural and phylogenetic analysis (Table 1). MTP is the smallest of the LLTP and contains little more sequence than the LLT module. In contrast, vertebrate apoB spans ~4500 aa, insect apoLp-II/I ~3350 aa, Vtg range from ~1350 (zebrafish Vtg1) to ~2600 aa (decapodan Vtg), CP ~1700 aa, M-177 ~1700 aa, ARP ~4350 aa, and MEP ~1460 aa (Figure 1 and 2). Homology among all disparate LLTP is limited to the LLT module, as evidenced by BLAST searches as well as multiple alignments. The single LLT module is always situated near the N-terminus, with the sole exception of Vtg from the water flea *D. magnalis* that has a 145 aa superoxide dismutase (SOD)-like domain, N-terminal to the LLT module. Most LLTP also contain a single

Table 1. LLTP selected for analysis. [1)] Sequence identifiers used in Figures 3 and 5; [2)] Species abbreviations, as used in Figures 1 and 2, are indicated in between parentheses; [3)] Sequence compiled from multiple sources (GenBank RefSeq); [4)] Sequence deposited but unpublished.

Sequence name [1)]	Species [2)]	Protein accession	Reference
apoB_Chicken	*Gallus gallus* (*Gg*)	XP_419979.1	Pruitt *et al.*, 2005 [3)]
apoB_Human	*Homo sapiens* (*Hs*)	AAA35549.1	Chen *et al.*, 1996
apoB_Zebrafish	*Danio rerio* (*Dr*)	XP_694827.1	Pruitt *et al.*, 2005 [3)]
apoLp_Fruitfly	*Drosophila melanogaster* (*Dm*)	AAC47284.1	Kutty *et al.*, 1996
apoLp_Locust	*Locusta migratoria* (*Lm*)	CAB51918.1	Bogerd *et al.*, 2000
ARP_Fruitfly	*Drosophila melanogaster* (*Dm*)	AAF52796.1	Adams *et al.*, 2000
CP_SignalCrayfish	*Pacifastacus leniusculus* (*Pl*)	AAD16454.1	Hall *et al.*, 1999
CP_Tigershrimp	*Penaeus monodon* (*Pm*)	AAF19002.1	Yeh *et al.*, 1999
M-177_Mite	*Euroglyphus maynei* (*Em*)	AAF14270.1	Epton *et al.*, 1999
MEP_Beetle	*Tenebrio molitor* (*Tm*)	BAB03250.1	Lee *et al.*, 2000
MEP_Fruitfly	*Drosophila melanogaster* (*Dm*)	AAM50129.1	Stapleton *et al.*, 2002
MTP_Chicken	*Gallus gallus* (*Gg*)	XP_420662.1	Pruitt *et al.*, 2005 [3)]
MTP_Fruitfly	*Drosophila melanogaster* (*Dm*)	AAF53946.2	Adams *et al.*, 2000
MTP_Human	*Homo sapiens* (*Hs*)	P55157	Sharp *et al.*, 1993; Shoulders *et al.*, 1993
MTP_Nematode	*Caenorhabditis elegans* (*Ce*)	AAR27937.1	Shibata et al., 2003
MTP_Zebrafish	*Danio rerio* (*Dr*)	CAD21747.1	Marza *et al.*, 2005
Vtg2_Chicken	*Gallus gallus* (*Gg*)	AAA49139.1	Nardelli *et al.*, 1987
Vtg_Lamprey	*Ichthyomyzon unicuspis* (*Iu*)	Q91062	Sharrock *et al.*, 1992
Vtg1_Zebrafish	*Danio rerio* (*Dr*)	NP_739573.1	Tong *et al.*, 2004
Vtg_Beetle	*Tenebrio molitor* (*Tm*)	AAU20328.2	Warr *et al.*, 2004 [4)]
Vtg1_Cockroach	*Periplaneta americana* (*Pa*)	Q9U8M0	Tufail *et al.*, 2000
Vtg_SilkMoth	*Bombyx mori* (*Bm*)	BAA06397.1	Yano *et al.*, 1994
Vtg_Crayfish	*Cherax quadricarinatus* (*Cq*)	AAG17936.1	Abdu *et al.*, 2002
Vtg_Shrimp	*Metapenaeus ensis* (*Me*)	AAM48287.1	Tsang *et al.*, 2003
Vtg_WaterFlea	*Daphnia magna* (*Dma*)	BAD05137.1	Kato *et al.*, 2004
Vtg1_Nematode	*Caenorhabditis elegans* (*Ce*)	AAB52675.1	*C. elegans* Sequencing Consortium, 1998
Vtg5_Nematode	*Caenorhabditis elegans* (*Ce*)	AAA83587.1	*C. elegans* Sequencing Consortium, 1998
Vtg6c_Nematode	*Caenorhabditis elegans* (*Ce*)	AAQ91901.1	*C. elegans* Sequencing Consortium, 1998
Vtg_PacificOyster	*Crassostrea gigas* (*Cg*)	BAC22716.1	Matsumoto *et al.*, 2003
Vtg_GalaxyCoral	*Galaxea fascicularis* (*Gf*)	BAD74020.1	Hayakawa *et al.*, 2004 [4)]

vWF-D module, with the exception of vertebrate apoB, MTP (all animals), MEP from the beetle *T. molitor*, Vtg from the oyster *C. gigas*, Vtg1 from zebrafish, and M-177, that is situated near the C-terminus. In the middle of the vWF-D module, the selected Vtg of vertebrates, nematodes and water flea have a CGLCG-like motif, reminiscent of the active

sites of proteins catalyzing disulfide bond formation. At this aligned position, insect Vtg and decapodan CP lack the N-terminal cysteine, whereas apoLp-II/I, ARP, and decapodan Vtg lack both cysteines (data not shown). A search for additional domains using the SMART/pfam database revealed a single additional motif of high significance (E<0.001), pfam06448, located from aa ~950 to aa ~1050, uniquely present in apoLp-II/I, ARP, decapodan Vtg, and apoB only. The presence of this pattern suggests that these vertebrate and arthropodan LLTP are relatively closely related.

The selected LLTP were analyzed for the presence of pronounced clusters enriched in either amphipathic α-helices or amphipathic β-strands, *i.e.* α and β clusters, using LOCATE prediction of amphipathic secondary structure (Figure 1). Comparison with the available lipovitellin crystal structure showed that, using the present cut-off values for calculated lipid affinity, only approximately half of the LOCATE predictions were accurate in the LLT module, as the size of the α-helical bundle was overestimated, whereas the β-strands in the βA-sheet were not recognized as amphipathic. Therefore, our discussion is limited to the more C-terminal pronounced clusters of predicted amphipathic secondary structure, the prediction of which is supported by additional data in apoB (Segrest *et al.*, 1994, 1998, 1999, 2001). For the included apoB, apoLp-II/I, ARP, as well as the Vtg from decapodan Crustacea and the oyster *C. gigas*, a pronounced alternating clustering of α and β clusters was predicted. The amphipathic clusters are distributed along these proteins as N-α-β-α-C, with the exception of apoB, for which an extended clustering conforming to N-α-β-α-β-α-C was predicted. For M-177, insect Vtg and *C. elegans* Vtg a less pronounced clustering according to N-α-β-C was observed, whereas clustering was less evident for CP, MEP, and vertebrate Vtg. As compared to other included LLTP family members, the amphipathic β-strand cluster(s) of the included apoB, apoLp-II/I, ARP, and decapodan Vtg are predicted to be relatively large, encompassing about ~1000 aa in decapodan Vtg, up to ~2400 aa in ARP, whereas the recognized β cluster of M-177 is smaller yet particularly enriched in amphipathic β-strands. In contrast, CP, MEP, MTP, and Vtg (except decapodan Vtg) either lack an evident β cluster or have a clearly smaller β cluster with less amphipathic β-strands.

Vertebrate Vtg most often contain multiple polyserine tracts, occupying a region of ~40 to ~210 aa (Figure 2). As noted below, several invertebrate Vtg contained polyserine tracts in the LLT module. Polyserine tracts were only identified in Vtg, and not in other LLTP. Both characterized decapodan CP contain a ~50 aa glutamine-rich region, with the shrimp CP additionally containing a polyglutamine tract near its C-terminus (Hall *et al.*, 1999; Yeh *et al.*, 1999). The features of selected LLTP, including the recognized modules, amphipathic clusters, established and putative post-translational cleavage sites, polyserine tracts and glutamine-rich regions, are summarized in Figure 2.

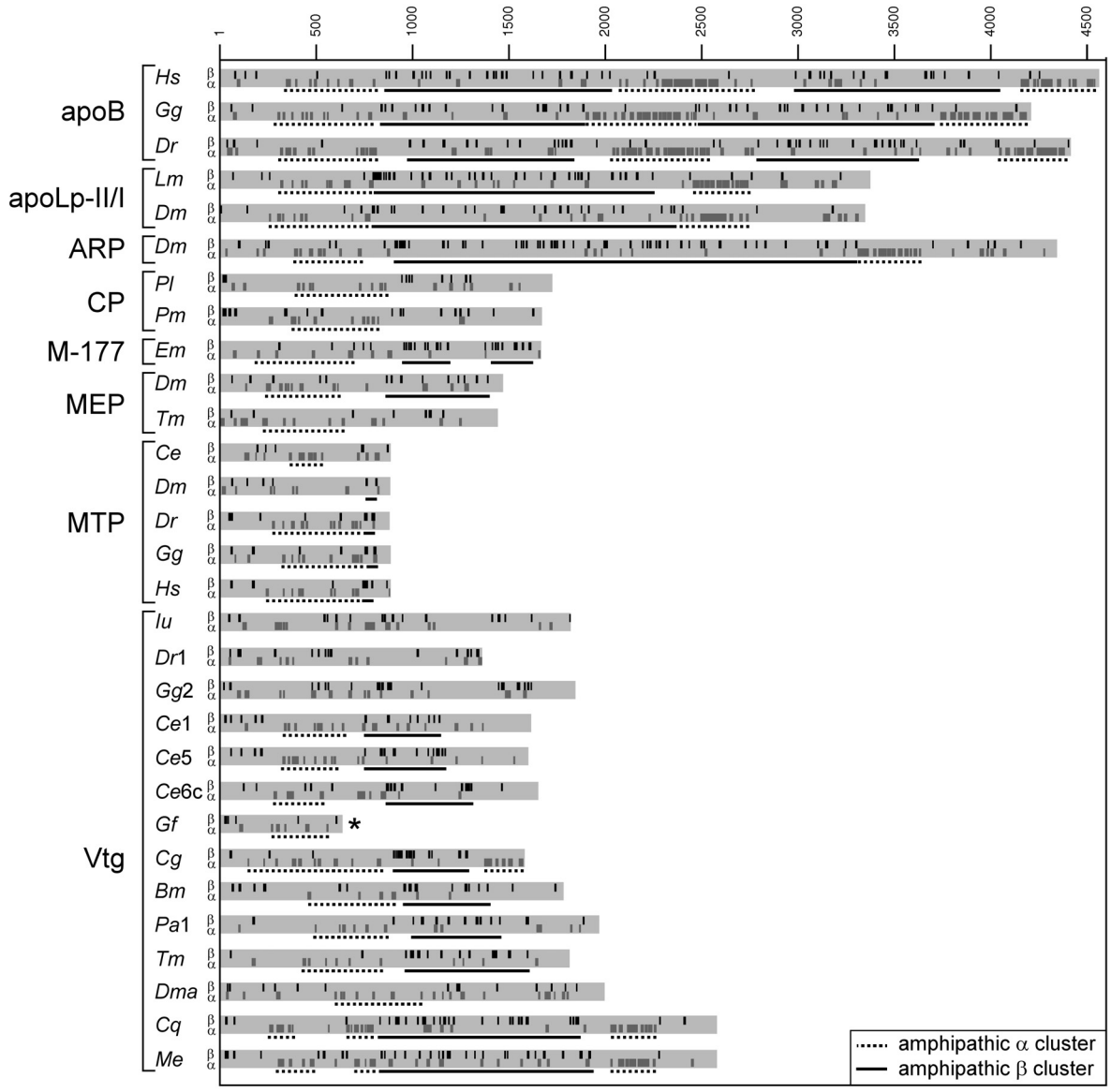


Figure 1. Clusters of amphipathic secondary structure in apoLp-II/I. Amphipathic α-helices and amphipathic β-strands in selected LLTP were predicted using LOCATE. Each LLTP is shown as a grey box with size proportionally to length, with the N- and C-terminus at the left and right sides, respectively. For each protein, the predicted α-helical or β-strand amphipathic stretches are indicated by grey and black boxes at their predicted amino acid positions (horizontal axis), provided their calculated lipid affinity is equal or above 10 and 7 kcal/mol, respectively. LLTP are named using the first letters of their scientific name, with the exception of the Vtg of *Daphnia magna* that is indicated with *Dma* to avoid confusion with sequences from *Drosophila melanogaster* (*Dm*) (Table 1). Optional numbers behind species abbreviations indicate the specific sequence, if multiple are known. Dotted and solid black lines below the panel of each LLTP indicate the clusters enriched in either amphipathic α-helices or amphipathic β-strands, respectively, as designated from the present results as well as additional sets of cut-off values for calculated lipid affinity (data not shown). An asterisk indicates that the analyzed sequence is not complete.

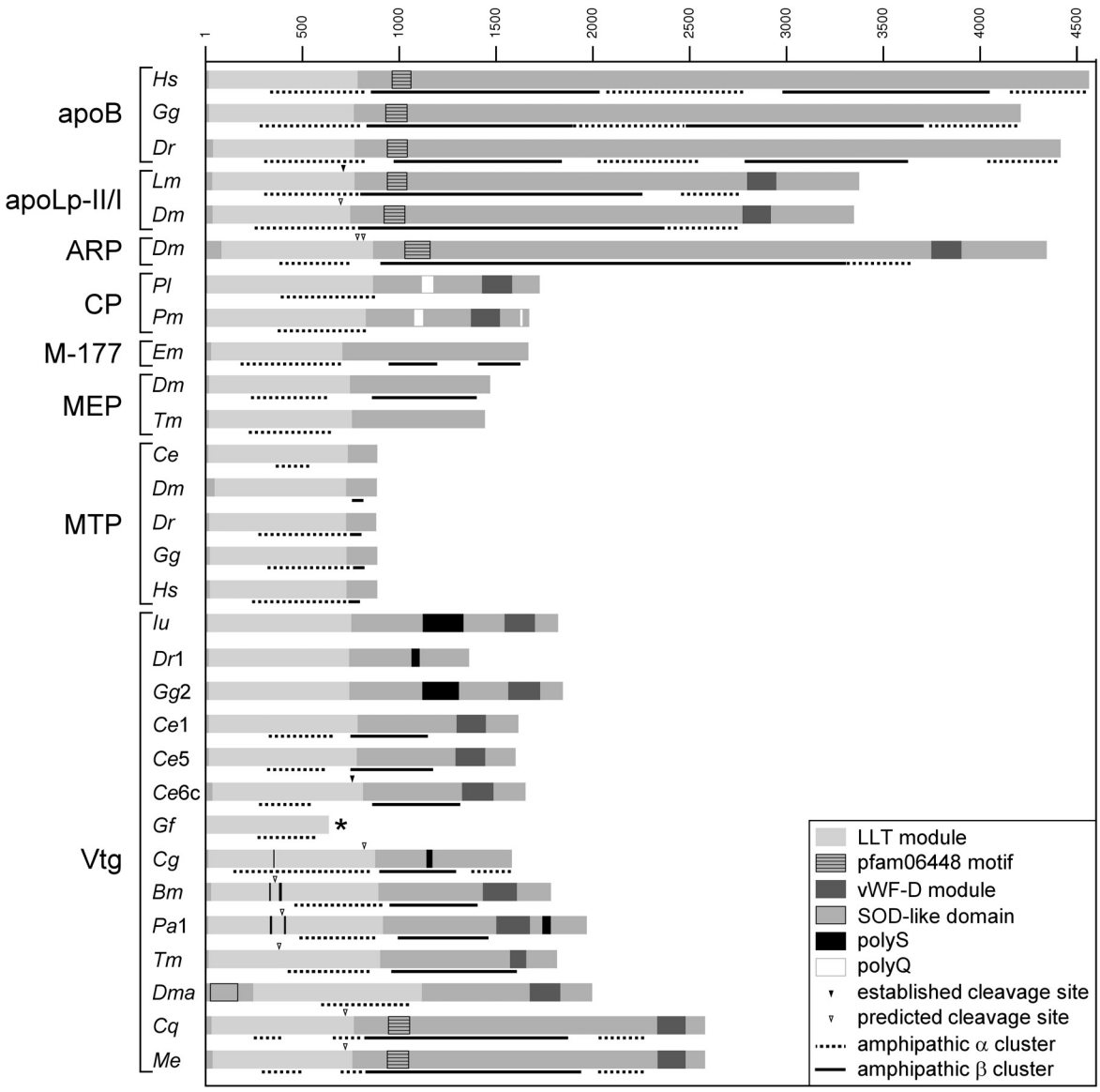


Figure 2. Modular architecture of selected LLTP. LLTP are shown with size proportional to length, with the N- and C-terminus at the left and right sides, respectively. Recognized domains are indicated by boxes at their respective position within the LLTP. The LLT module, delimited by the presently recognized conserved segments (Figure 3), is indicated by a light grey box, the pfam06448 motif by a box with parallel horizontal lines, the vWF-D module by a dark grey box, and the SOD-like domain in the Vtg of *Daphnia magna* (*Dma*) by a black rectangle. Regions rich in polyserine tracts are indicated by black boxes, whereas the polyglutamine regions in CP are indicated by white boxes. Predicted clusters enriched in amphipathic α-helices or amphipathic β-sheets are indicated by dotted and solid black lines below each LLTP, respectively. Evidenced and predicted sites of post-translational cleavage during biosynthesis are indicated by filled and open arrowheads, respectively. An asterisk indicates that the analyzed sequence is not complete. LLTP are named as in Figure 1.

Architecture of the LLT module

The N-terminal protein sequences from selected LLTP (Table 1) were aligned to allow for analysis of the conserved (structural) regions, molecular diversity and phylogenetic relation of their LLT modules. Sequence divergence is high and about 60% of the total LLT sequence covered cannot be aligned confidently among all included LLTP. However, 13 sequence blocks of contiguous conserved sequence could be detected, in total including 337 to 350 aa of the total LLT module with on average ~40% similarity among selected LLTP (Figure 3). These conserved segments largely correspond to the conserved N1 to N5 and N7 to N20 sequence motifs, previously recognized by Babin and co-authors (1999). The additional conserved motifs observed in that study, N6, N21, and N22, could not be aligned unambiguously in the present set of LLTP due to the relatively high sequence divergence of the MTP from fruitfly and *C. elegans*.

Several remarkable features were noted in between the conserved segments of some selected LLTP. The number of residues in between segments 4 and 5 (corresponding to N4 and N5 in Babin *et al.*, 1999) is far larger in both decapodan CP than in others (79 and 91 aa *vs.* 3 to 28 aa), and the additional residues are marked by the presence of several repeated lysine-rich motifs (Hall *et al.*, 1999; Yeh *et al.*, 1999), *i.e.* five S-K-T-S motifs in *P. monodon* (shrimp) CP and three T-K-T-T-G motifs in the *P. leniusculus* (crayfish) CP. In between segments 5 and 6 (corresponding to N5 and N7 in Babin *et al.*, 1999), the selected insect Vtg contain putative cleavage sites, corresponding to the consensus substrate sequence for furin, R-X-K/R-R. Strikingly, these cleavage sites are located in between two nearby polyserine tracts in cockroach and silk moth Vtg. Established and predicted cleavage sites are also located between the C-terminal segments 12 and 13 (corresponding to N19 and N20 in Babin *et al.*, 1999) in all included apoLp-II/I as well as the Vtg of decapodan Crustacea, Pacific oyster, and *C. elegans*. In between segments 9 and 10 (corresponding to N11 and N12 in Babin *et al.*, 1999), Vtg of water flea contains an additional stretch of ~60 aa, relatively enriched in lysines (8 within 19 aa, organized as 4 doublets).

The conserved sequence segments as well as above-mentioned remarkable variable regions in LLT modules were mapped onto the available crystal structure of the lamprey lipovitellin monomer (PDB accession: 1LSH; Thompson and Banaszak, 2002) that contains the LLT module of lamprey Vtg (Figure 4). Conserved sequence blocks 1 to 5 are situated in the barrel-like βC-sheet. A relatively highly conserved motif N-L/I/V/M-K/R-K/R-G-L/I/V was noted in conserved segment 3, only absent from *C. elegans* MTP, that corresponds to the α-helix inside the barrel-like βC-sheet. This motif appears to interact with another highly-conserved motif P-S/T-X-G-L/V/I/M-P, located in a loop extending from the βB-sheet into the barrel-like βC-sheet. Latter motif is situated in the N21 conserved sequence motif (Babin *et al.*, 1999) and is present in all LLTP, except for *D. melanogaster* and *C. elegans* MTP. The presently observed conserved segments 6 to 10

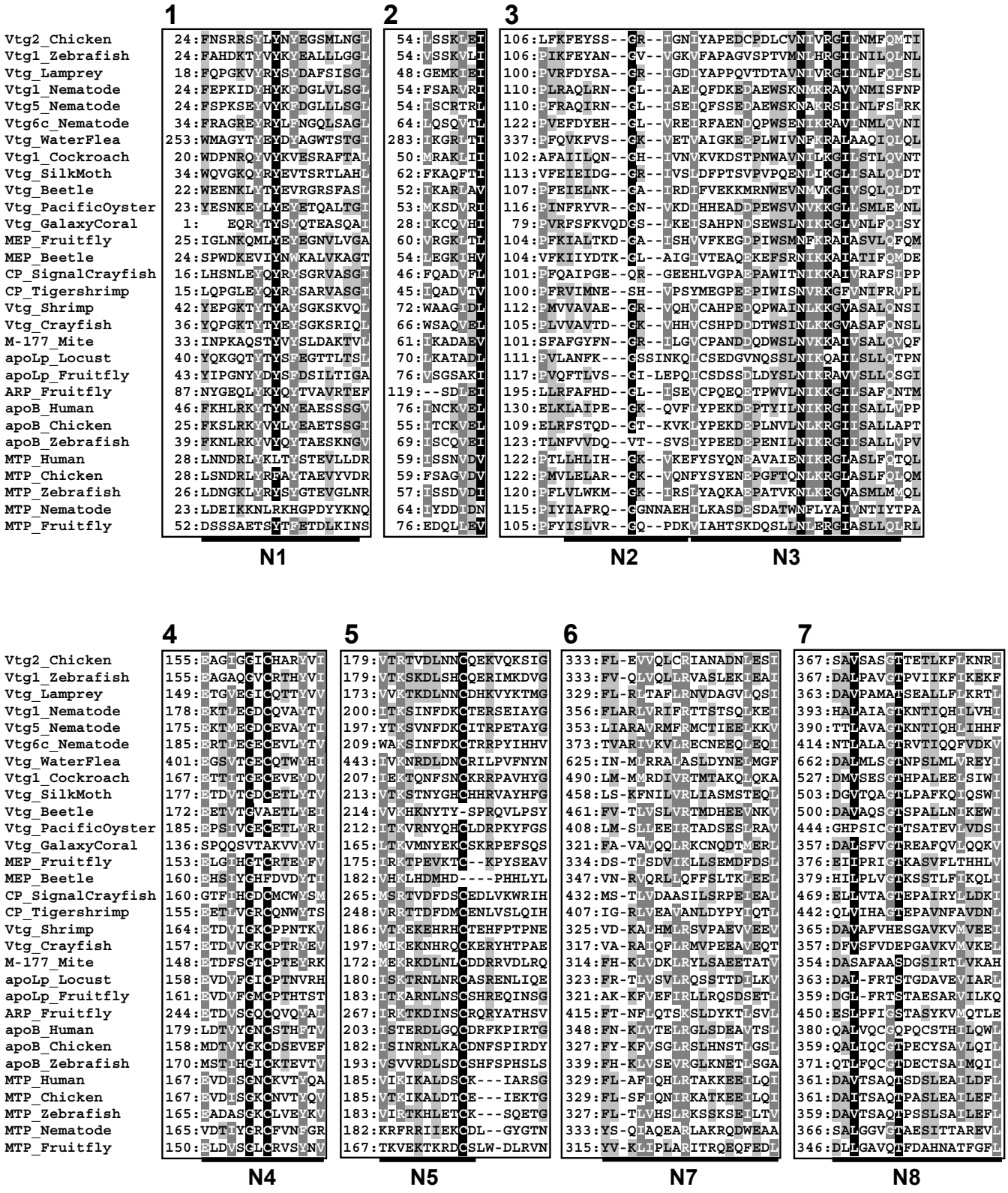
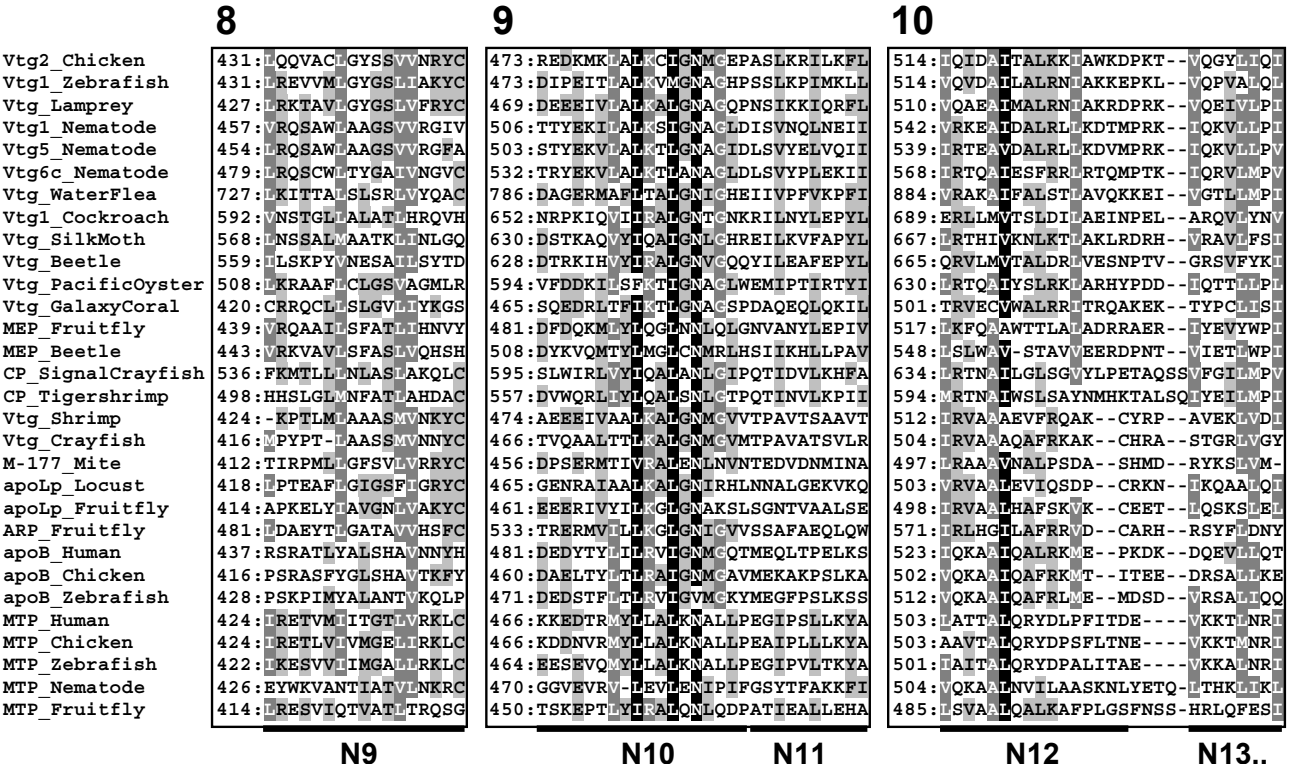
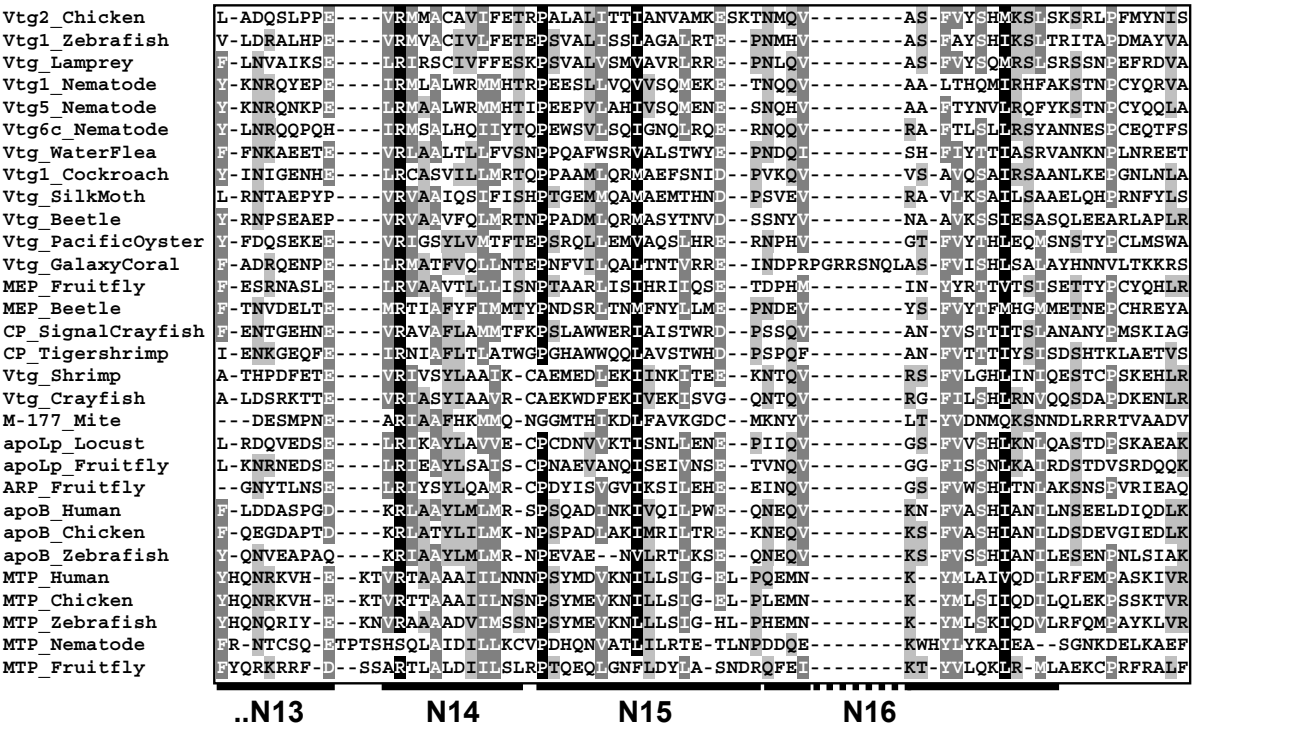


Figure 3. Alignment of the LLT modules from selected LLTP. LLTP are named according to the sequence names in Table 1. The recognized conserved segments are shown in boxes, with their identification number indicated above each box. Thick lines at the bottom of blocks indicate the location of the conserved LLT module motifs recognized by Babin *et al.* (1999). Numbers at the left of conserved sequence blocks indicate the N-terminal starting position of the conserved sequence blocks for each sequence. Amino acid residues →





identical or considered conserved in over 90%, in between 90% and 65%, or in between 65% and 40% of included sequences are shaded in black, dark gray and light gray, respectively. The allowed conservative substitutions were defined as follows: A, G; S, T; E, D; R, K, H; Q, N; V, I, L, M; Y, F; W; P; and C. Dashes indicate gaps that were introduced in conserved blocks to optimize alignments, whereas blank spaces are introduced if the sequence is incomplete. **(figure continued on next page)**

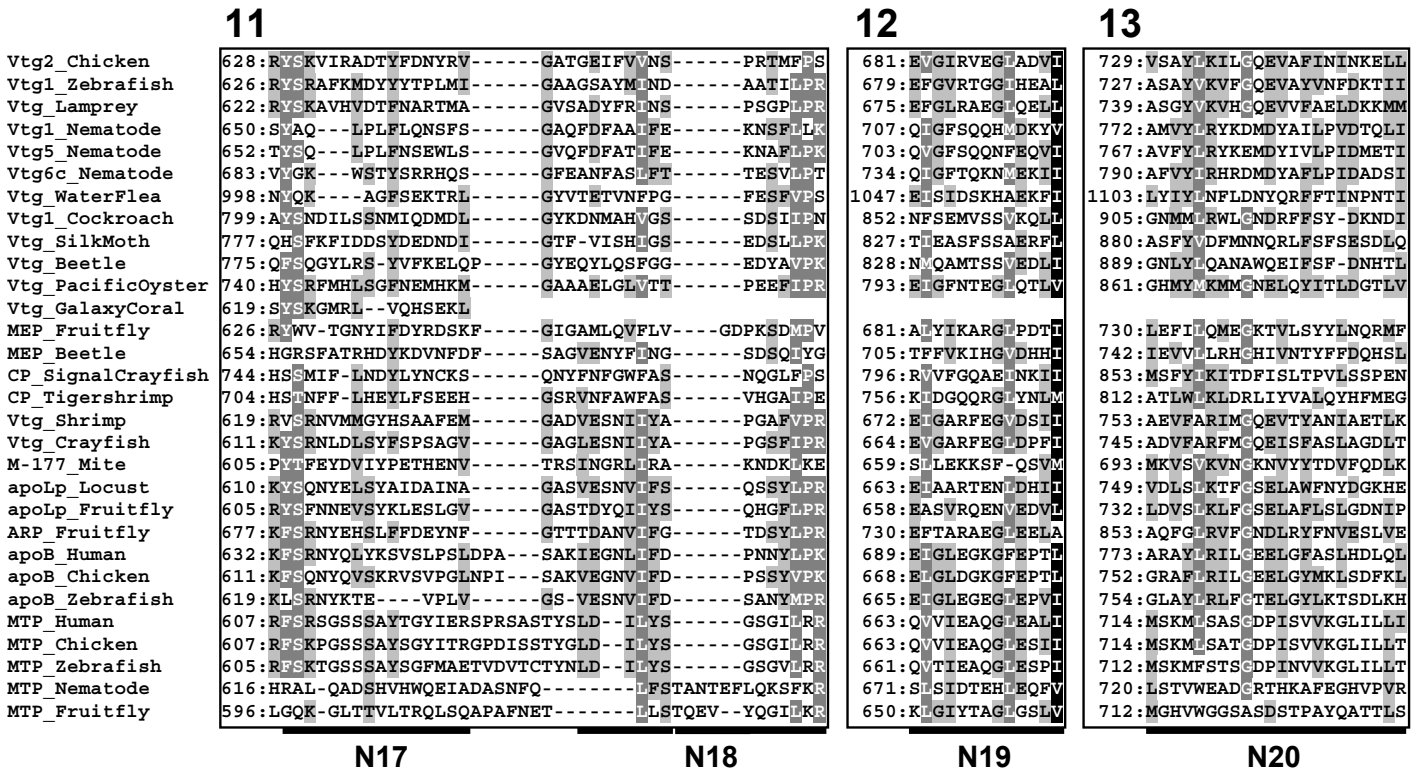


Figure 3. (continued from previous pages)

correspond mostly to α-helices in the α-helix bundle region, whereas segments 11 to 13 correspond to the βA-sheet of the lipid-binding cavity. Considering the βB-sheet, sequence conservation was noted only in the motifs corresponding to N21 and N22 in Babin *et al.* (1999) that are not present in the MTP of fruitfly and *C. elegans*. Several of remarkable LLT module inserts appear to be located around the barrel-like βC-sheet, *i.e.* at its top, as for the repeated lysine-rich motifs of both decapodan CP, or on its backside, as for the furin-like cleavage motifs and polyserine tracts in insect Vtg (Figure 4b). In addition, the lysine-rich insert in the Vtg of water flea is situated within a loop between two α-helices in the α-helical bundle. As noted previously for locust apoLp-II/I (Chapter 2), the post-translational cleavage site of insect apoLp-II/I as well as the Vtg from several invertebrates (decapodan Crustacea, Pacific oyster and *C. elegans*) is located in a region of unknown structure, in between two β-strands of the βA-sheet (Figure 4).

Phylogenetic analysis of LLTP

Phylogenetic analysis on the conserved sequence blocks in the LLT module revealed topologically similar unrooted trees, using ML as well as Bayesian analysis methods (Figure 5). Three distinct subfamilies of LLTP can be discerned in the present unrooted tree. The first subfamily is constituted by vertebrate and invertebrate MTP and is strongly supported by Bayesian as well as ML bootstrap analyses (100% and 99% support,

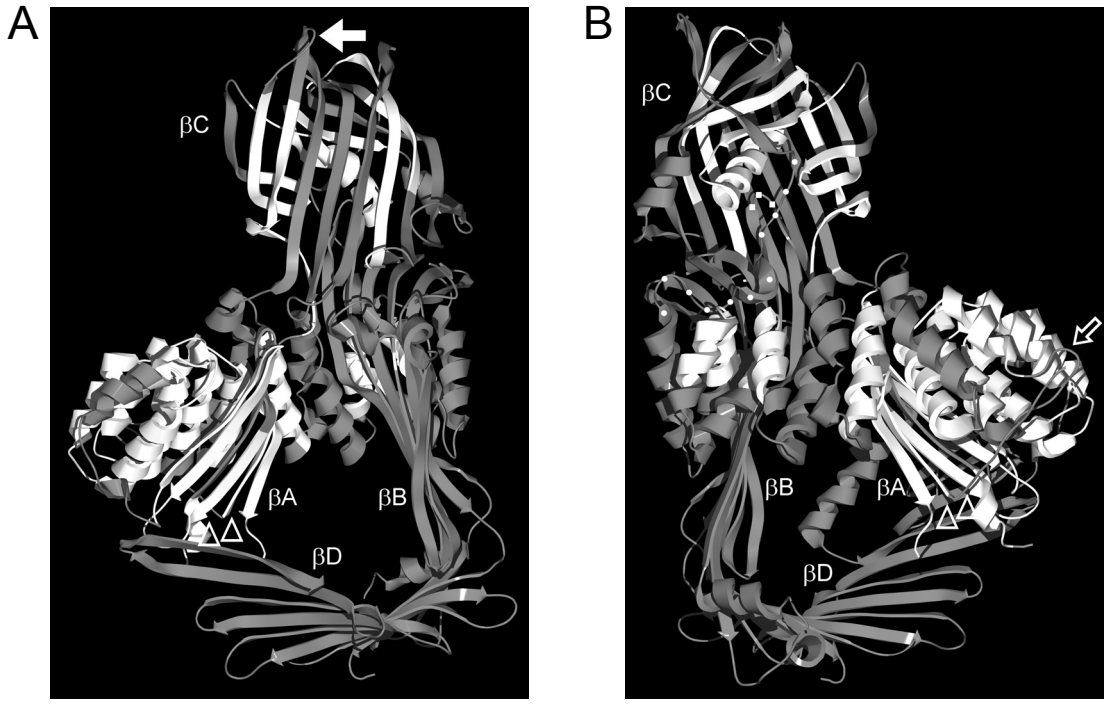


Figure 4. Mapping of LLT module sequence variation onto the lamprey lipovitellin structure. Ribbon representation of the lamprey lipovitellin crystal structure that includes the complete LLT module. Conserved regions across LLTP are shown in white, and the rest in grey. **A**: Front view, several structures outside of the LLT module obstructing the view on the lipid-binding cavity (C-terminal from aa 1456) have been left out; **B**: Back view. Regions corresponding to conserved sequence motifs are shown in white and others in grey. The three β-sheets present in the LLT module are indicated by βA, βB, and βC. Sequences C-terminal to the LLT module encode part of the βB and the whole βD sheet. The closed arrow in the front view (A) indicates the loop in which decapodan CP have an insert of repeated lysine-rich motifs. In the back view (B), an open arrow indicates the loop in which the Vtg of water flea contains an additional stretch of ~60 aa, relatively enriched in lysines (8 within 19 aa, organized as 4 doublets), white circles indicate the region that contains the two polyserine tracts with intermittent cleavage site in cockroach and silk moth Vtg, and white squares indicate the location of the conserved motif in a loop extending from the βB-sheet into the barrel-like βC-sheet. Open arrowheads point to the β-strands of the βA-sheet that are connected by a loop of unknown structure that contains a consensus substrate sequence for cleavage by furin in apoLp-II/I, ARP, decapod Vtg and several other invertebrate Vtg.

respectively). *D. melanogaster* and *C. elegans* MTP are grouped together, and have particularly long branch lengths, in line with the sequence divergence noted in the LLT module alignment (Figure 3).

The second subfamily, referred to as apoB-like LLTP, includes vertebrate apoB, insect apoLp-II/I, fruitfly ARP, mite M-177, and Vtg from decapodan Crustacea. The grouping of these sequences together, apart from others, received 95.1% and 51% support in the Bayesian and ML analyses, respectively. Phylogenetic analysis of the aligned vWF-D modules (in total only ~100 aa alignable sequence) revealed the grouping of decapodan Vtg with insect apoLp-II/I and ARP, apart from an unresolved group of other LLTP with a

vWF-D module, further supporting that decapodan Vtg are apoB-like LLTP (data not shown). Although the fruitfly ARP is clearly related to apoLp-II/I, phylogenetic support for its grouping with apoLp-II/I rather than other apoB-like LLTP is relatively poor, highlighting the molecular diversification of ARP as an apoB-like LLTP. Bayesian and ML analyses do not agree on the placement of M-177. Bayesian analysis provides 100% support for the grouping of M-177 together with apoLp-II/I, ARP and decapodan Vtg, however, the exact position of M-177 is unresolved; ML analysis supports this grouping with 53%, whereas there is 36% support for a basal position of M-177 with the MTPs. Given the particularly long branch lengths of M-177 and the MTP of *D. melanogaster* and *C. elegans*, we note that this relatively high alternative support could reflect long-branch attraction.

The third subfamily, referred to as Vtg-like LLTP, includes vertebrate as well as invertebrate Vtg (except for Vtg from decapodan Crustacea), decapodan CP and insect MEP. The grouping of these sequences receives 98.2% support in Bayesian analysis and 57% in ML bootstraps.

Support for the grouping of any of the LLTP, except M-177, to members of another subfamily is very low in Bayesian and ML analyses (data not shown), corroborating the validity of the proposed division of LLTP along subfamilies. Thus, three subfamilies of LLTP are recognized in the obtained unrooted phylogenetic tree, named MTP, apoB-like LLTP and Vtg-like LLTP.

Discussion

Subfamilies of LLTP

The LLTP family consists of homologous lipid-binding proteins of high physiological and pathological importance, defined by the presence of an N-terminal conserved region, the LLT module. We presently investigated the diversity of LLTP to gain additional understanding of the molecular evolution, origin and adaptation of this protein family. Investigation of annotated genomes showed that distinct animal taxa have specific subsets of LLTP. Most insects have genes encoding MTP, apoLp-II/I, most often one or several Vtg, as well as two newly recognized family members, *i.e.* melanin-engaging protein (MEP) and apoLp-II/I-related protein (ARP). Analysis of the human, chicken and zebrafish genomes demonstrates that vertebrates have MTP, apoB and, except mammals, additionally encode one or more Vtg. In contrast, the nematode *C. elegans* has no apoB-like protein but encodes MTP as well as seven Vtg. As yet, two distinct LLTP have been identified in distinct species of decapodan Crustacea, the Vtg-like CP as well as a Vtg that actually is an apoB-like LLTP. Given the relative close relation among these species, it

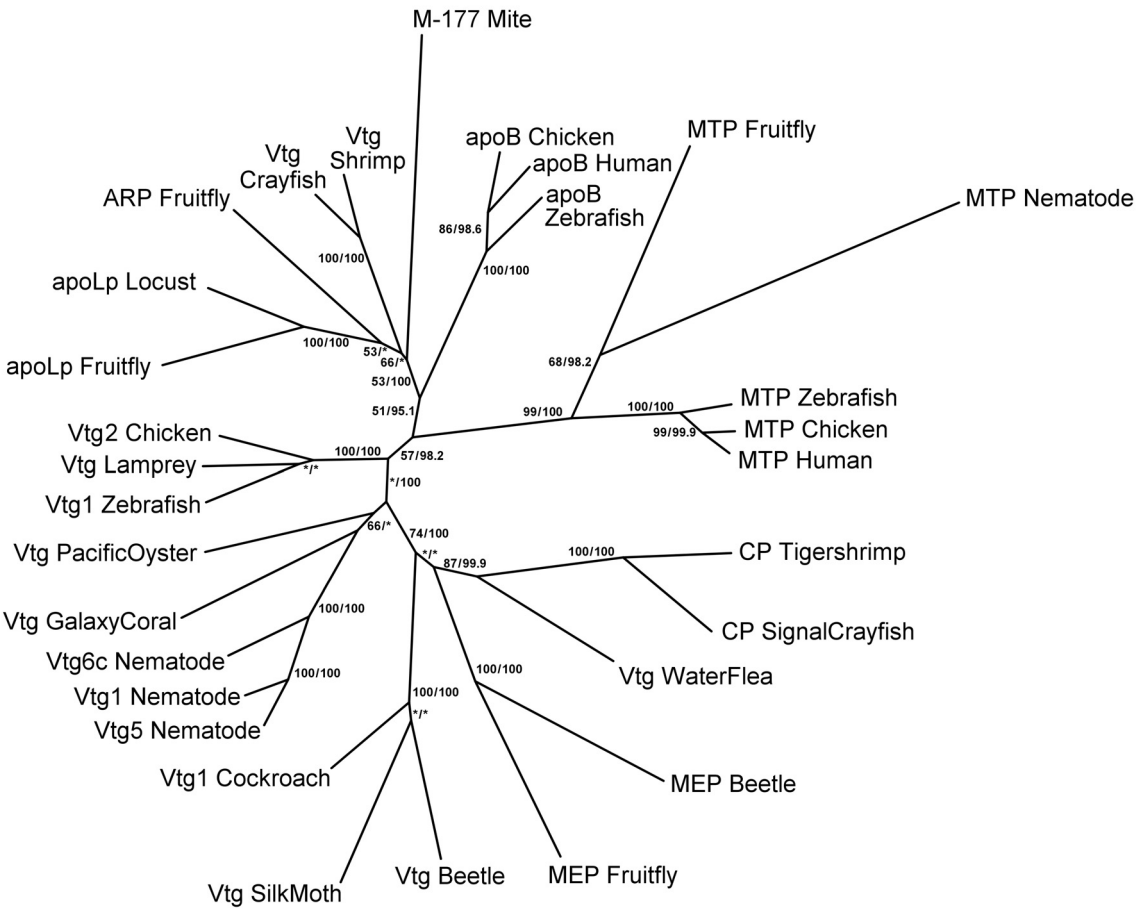


Figure 5. Unrooted maximum likelihood tree of LLTP sequences. The tree is based on conserved segments of the LLT module using the WAG + I + Γ_4 model of sequence evolution. Numbers indicate ML bootstrap and Bayesian posterior probability values (left and right of slash, respectively). Only support values equal or above 50% for ML and 95% for Bayesian analyses are shown, whereas nodes with lower support are indicated by an asterisk.

appears very likely that decapodan Crustacea have both CP and Vtg. Thus, distinct animal taxa have specific, yet mostly overlapping subsets of LLTP.

Phylogenetic analysis on conserved segments in the LLT module provides insight in the evolution of LLTP as it reveals three major subfamilies of LLTP: (1) vertebrate and invertebrate MTP, (2) vertebrate apoB, insect apoLp-II/I, insect apoLp-II/I-related protein, mite M-177 allergen and decapodan Vtg, and (3) Vtg (excluding decapodan Vtg), CP and MEP. This grouping likely extends to the identified LLTP not included in the present phylogenetic analysis, as these have a relatively high similarity to at least one of the included LLTP. From here on, we refer to these three subfamilies as MTP, apoB-like LLTP, and Vtg-like LLTP, respectively.

Within the MTP subfamily of LLTP, the nematode and fruitfly sequences are grouped together apart from the vertebrate sequences, as in the Vtg-like LLTP. The MTP

of *D. melanogaster* and particularly of *C. elegans* appear to have diverged strongly, as also noted in the alignment. In nematodes, insects as well as vertebrates, MTP stimulates the biosynthesis of other LLTP (Shibata *et al.*, 2003; Shelness *et al.*, 2005; Shelness and Ledford, 2005; Chapter 3). Since the *C. elegans* MTP has only one type of LLTP substrate, the relatively closely related Vtg, this divergence may reflect the absence of the selective pressure induced by multiple distinct substrates, as is the case in insects as well as vertebrates that have multiple distinct LLTP substrates, *i.e.* Vtg-like and apoB-like LLTP.

The subfamily of Vtg-like LLTP is highly diverse. Previous phylogenetic analyses of Vtg differed regarding the relation among nematode, insect, and vertebrate Vtg. Several analyses based on the conserved N-terminal regions of Vtg revealed a closer relation between the Vtg from nematodes and insects as compared to Vtg from vertebrates (Lim *et al.*, 2001; Kato *et al.*, 2004). In contrast, other analyses suggest a closer relation between the Vtg from nematodes and vertebrates as compared to Vtg from insects (Trewitt *et al.*, 1992; Winter *et al.*, 1996; Chen *et al.*, 1997; Sappington and Raikhel, 1998a). However, most of the latter analyses may not be interpreted correctly, since the Vtg trees were unrooted (Chen *et al.*, 1997), "rooted" using an insect Vtg rather than a non-Vtg LLTP (Winter *et al.*, 1996), or presented and interpreted without taking the included apoB outgroup into account (Chen *et al.*, 1997; Sappington and Raikhel, 1998a). The present phylogenetic analysis suggests a closer relation between the Vtg from nematodes and insects as compared to Vtg from vertebrates, in accordance with a present view on animal phylogeny (Maddison and Schulz, 2004). In addition to multiple related Vtg, insects also have another Vtg-like LLTP, MEP, that likely is a paralog of an insect Vtg. In decapodan Crustacea, however, CP is the only Vtg-like LLTP identified at present, as the protein named Vtg in this taxon is found to be an apoB-like LLTP.

The grouping of vertebrate apoB, insect apoLp-II/I, insect ARP, and decapodan Vtg in a subfamily is additionally supported by the recognition of a homologous region in these sequences only, the pfam06448 motif (Marchler-Bauer *et al.*, 2005), located just C-terminal to the LLT module. Moreover, members of this subfamily have a N-terminal distribution of pronounced amphipathic clusters with a relatively long β cluster and additional α cluster, according to N-α-β-α-C. Phylogenetic analysis suggests that mite M-177 is also a member of this apoB-like LLTP subfamily, although it lacks the pfam06448 motif, extended β cluster and additional α cluster, present in other apoB-like LLTP. Additional sequences of LLTP from arachnids may help define the context of its divergence. The present grouping of vertebrate apoB and insect apoLp-II/I could not be demonstrated convincingly previously (Babin *et al.*, 1999; Mann *et al.*, 1999), likely due to the unavailability of the presently applied phylogenetic methods. Importantly, the Vtg of decapodan Crustacea are not directly related to the Vtg genes of other taxons, as assumed previously (Tsutsui *et al.*, 2000; Abdu *et al.*, 2002; Okuno *et al.*, 2002; Avarre *et al.*, 2003; Tsang *et al.*, 2003; Kung *et al.*, 2004; Tsutsui *et al.*, 2004; Mak *et al.*, 2005;

Raviv *et al.*, 2006), but are apoB-like LLTP. This observation explains previous phylogenetic analyses on Vtg only (Lim *et al.*, 2001; Tsang *et al.*, 2003; Kato *et al.*, 2004; Donnell, 2004), in which decapodan Vtg did not group with other invertebrate Vtg.

Origin of modules and subfamilies of LLTP

The MTP, apoB-like and Vtg-like subfamilies of LLTP appear to have arisen prior to the evolution of Bilateria, given their distribution in distinct phyla. The arisal of the LLT module appears to be the hallmark event in the origin of the complete family of LLTP, as it provides the basal structure for the binding of multiple lipid molecules. Based on the absence of sequences similar to the LLT module in taxa other than animals, the LLT module likely is a Metazoan innovation, that arose at least prior to the divergence of the Cnidaria and bilaterian phyla. Thus, the evolution of the LLT module may coincide with the evolution of animal multicellularity, a condition that provoked the need for intercellular lipid transport. Previously, the common ancestor to the present LLTP has been suggested to function in vitellogenesis, as this ancient process is essential to reproduction in even the oldest animal phyla. More recently, however, MTP has been proposed as the predecessor to other LLTP, in view of its importance in the biosynthesis of other LLTP (Shibata *et al.*, 2003; Shelness *et al.*, 2005; Shelness and Ledford, 2005; Chapter 3). Nonetheless, it cannot be excluded that the LLT module evolved as part of a larger multidomain protein, with a function unrelated to any of the functions presently ascribed to LLTP. Thus, the major LLTP subfamilies likely arose in an ancestor of the Bilateria, but the early history of the LLT module remains shrouded.

Recent preliminary apoB sequences from the emerging sea urchin genome (accessions XP_785757 and XP_785809) do suggest a common origin for the current Vtg-like and apoB-like LLTP. The sea urchin apoB contains an LLT module and is predicted to have a pentapartite amphipathic structure similar to vertebrate apoB (data not shown), but additionally has a C-terminal vWF-D module. This poses the possibility that the ancestral vertebrate apoB lost its vWF-D module. Accordingly, the presence of a vWF-D module appears to be the ancestral state for apoB-like LLTP, suggesting that both apoB-like and Vtg-like LLTP stem from a common ancestor that consisted of a LLT module as well as a vWF-D module. This appears to be a more parsimonious possibility than the independent gain of the vWF-D module in apoB-like and Vtg-like LLTP. In addition to apoB, several Vtg-like LLTP have lost the vWF-D module, *e.g.* Vtg1 in zebrafish.

The organization of amphipathic clusters in LLTP has evidently been modified, as exemplified by the apoB-like LLTP, in which the number, size and organization of these clusters clearly vary. As suggested previously (Chapter 3), the amphipathic clusters in vertebrate apoB may have expanded from an ancestral N-α-β-α-C organization to the pentapartite N-α-β-α-β-α–C clustering, by the duplication of a β and α cluster in the ancestral vertebrate apoB gene. As compared to apoLp-II/I, the amphipathic β cluster

region of ARP appears to have expanded, resulting in the larger size of ARP. However, reconstruction of these modifications is prevented by the scarce sequence similarity within and among the involved regions of different LLTP.

Molecular diversity in LLTP

The subfamilies of LLTP have common and specific features in their modular organization that likely relate to functional aspects (Table 2). Comparison of various LLTP reveals that the LLT module is virtually always located at the N-terminus. The only exception presently found is the Vtg of *Daphnia magnalis* (water flea), in which the LLT module is preceded N-terminally by a SOD-like domain (Kato *et al.*, 2004). The fusion of this domain to Vtg has been suggested to play a role in transport of Cu(II) or in the immediate detoxification of superoxides resulting from Vtg metabolism (Kato *et al.*, 2004). The conservation of the N-terminal position of the LLT module in all other LLTP likely relates to its importance to lipoprotein assembly, as it folds into a basal lipid-binding structure that enables the proper co-translational folding and lipid association of further C-terminal amphipathic regions, as demonstrated for apoB (for reviews see Segrest *et al.*, 2001; Shelness and Sellers, 2001).

Conserved sequences in the LLT module suggest that all LLTP contain the major structural regions of the LLT module, with the possible exception of the *D. melanogaster* and *C. elegans* MTP sequences, as these LLTP display no sequence similarity associated to the βB-sheet. In the periphery of the βB-sheet, these LLTP also lack the highly conserved motif, located in a loop extending from the βB-sheet into the barrel-like βC-sheet, that appears to interact with another highly conserved α-helix within the barrel-like βC-sheet (Figure 4). Previously, *D. melanogaster* and *A. mellifera* MTP were already noted to diverge from vertebrate MTP by the absence of amphipathy in the helix A, at the boundary of the βB-sheet (Read *et al.*, 2000; Sellers *et al.*, 2003). These findings suggest that a large part of the βB-sheet and peripheral structure may not be present in *D. melanogaster* and *C. elegans* MTP. Given the role of these structures in constituting the lipid-binding cavity and mediating lipid transfer, their absence may have profound functional implications. Strikingly, *D. melanogaster* MTP nonetheless stimulated apoLp-II/I lipoprotein biogenesis *in vitro* (Chapter 3).

Within the non-conserved segments of their LLT module, several LLTP contain conspicuous insertions that appear to be located mostly in short loops in between α-helices and/or β-strands (Figure 4). In line with possible functional interactions, these insertions are located on the outer surface of the LLT module. For example, the repeated lysine-rich motifs in CP are located on top of the barrel-like βC-sheet, a location that may allow for the suggested role of these motifs in the crosslinking of CP by transglutaminase upon hemolymph coagulation (Hall *et al.*, 1999; Yeh *et al.*, 1999).

LLTP subfamilies differ to a great extent in the presence of amphipathic clusters (Figure 1). ApoB-like LLTP were predicted to contain one or two extended clusters enriched in amphipathic β-strands. The (partial) presence of such a β cluster dramatically increases the lipid binding of truncated apoB and apoLp-II/I (Chapter 3; for reviews see Hevonoja *et al.*, 2000; Segrest *et al.*, 2001). In apoB, the two β clusters have been predicted to form nearly continuous amphipathic β-sheets of sufficiently high lipid affinity to nearly irreversibly associate with lipids (Segrest *et al.*, 2001; Wang and Small, 2004). Thus, the β clusters in apoB-like LLTP may function to enhance their lipid-binding capacity, in accordance with the role of these LLTP in lipid distribution. Accordingly, the suspected role of MTP in the transfer of lipids to other LLTP may relate to its relatively limited amount of amphipathic β-strands. Vtg-like LLTP are also predicted to contain amphipathic β-strands, yet often in smaller or less clearly defined clusters (Figure 1; Segrest *et al.*, 1999). As compared to Vtg-like LLTP, the apoB-like decapodan Vtg contain an extended amphipathic β cluster, suggesting that they may have a larger lipid-binding capacity than any member of the Vtg-like LLTP.

In addition to a N-terminal α cluster, apoB-like LLTP also contain one or two more pronounced C-terminal α clusters of higher calculated lipid affinity. The amphipathic α-helices of the α clusters may reversibly associate or dissociate from the lipid core upon changes in lipid content (Wang *et al.*, 2003), as occur *in vivo* in lipid transport by lipoproteins constituted by apoB-like LLTP, *e.g.* the insect lipophorin and the vertebrate lipoproteins containing apoB. As compared to vertebrate apoB, insect apoLp-II/I contains only one such α cluster that also is of relatively small size. Therefore, the exchangeable apolipoprotein apoLp-III, that is constituted of amphipathic α-helices and enables additional lipid binding by insect lipoprotein, has been proposed to compensate for the relative small amount of amphipathic α-helices present in apoLp-II/I, as compared to apoB (Chapter 3). From this point of view, the absence of a pronounced C-terminal α cluster in nearly all Vtg-like LLTP accords with their limited dynamics in lipid content, as expected from their functions in yolk supply (Vtg), coagulation (CP) and immunity (MEP) (Table 2).

The vWF-D module, present in many apoB-like LLTP and most Vtg-like LLTP, is always located near the C-terminus. Unlike any LLTP, mammalian secretory mucin and von Willebrand factor contain multiple vWF-D modules that allow their multimerization (Journet *et al.*, 1993; Perez-Vilar and Hill, 1999). The functional significance of the single vWF-D module in LLTP is thus far unknown. Interestingly, the vWF-D module of several LLTP contains a complete CGLCG-like motif that has been been suggested to catalyze disulfide bonds during protein folding (Perez-Vilar and Hill, 1999). A similar motif of vicinal cysteines is present in protein disulfide isomerase, the binding partner that enables MTP function in mammals (for review see Shoulders and Shelness, 2005).

Unlike other LLTP, several Vtg-like LLTP also contain polyserine tracts. Although their physiological role is not clear, the serines are heavily phosphorylated, likely resulting in the binding of calcium (Montorzi *et al.*, 1995; Sappington and Raikhel, 1998a). Strikingly, the regions enriched in polyserine tracts are particularly large in most vertebrate Vtg, suggesting a taxon-specific role. Accordingly, we speculate that the extensive polyserine regions in vertebrate Vtg represent a molecular adaptation to enhance calcium deposition to the developing oocyte, to enable skeletal formation in the future embryo.

In conclusion, much of the diversity in the molecular organization of LLTP, within and between subfamilies, may be related to specific functional aspects that remain to be established experimentally.

Functional diversification of LLTP subfamilies

The ancient metazoan origin of LLTP subfamilies implies a rich history of LLTP gene duplications. In general, copies of genes are lost due to degenerative mutations, whereas only a minority is preserved on the long term, either by subfunctionalization, *i.e.* the gene copies partition/complement the function of the ancestral gene, or neofunctionalization, *i.e.* acquirement of a novel function by one of the copies (Lynch and Conery, 2000). The occurrence of frequent gene duplication in LLTP is clearly illustrated by the large variation in the number of Vtg genes between taxa and even within closely related species, *e.g.* seven Vtg genes are present in the zebrafish *D. rerio* (Wang *et al.*, 2005), two to 30 within different species of salmonids (Buisine *et al.*, 2002), four in the amphibian *Xenopus laevis* (Wahli *et al.*, 1979), and six in the nematode *C. elegans* (Blumenthal *et al.*, 1984). Vtg copies have been found to originate from genome duplication as well as local gene amplification, resulting in tandem repeats of Vtg genes (Buisine *et al.*, 2002; Wang *et al.*, 2005). Several of the multiple Vtg genes are highly similar, *e.g.* vitellogenins 3, 4 and 5 in *C. elegans*, suggesting either a recent origin, strong purifying selection for sequence, or concerted evolution. Such multiple Vtg genes may be maintained by subfunctionalization, as egg production and fertility can depend on the number of yolk-protein gene copies (Bownes *et al.*, 1991). Conversely, Vtg genes can be lost entirely when they become redundant, as in the higher Diptera whose ancestor reverted to a lipoprotein lipase-related protein as the major yolk protein (Sappington, 2002), and in the ancestral mammal that abandoned yolk accumulation in the oocyte altogether.

The apoB-like LLTP and MTP would appear to be as prone to duplication as the Vtg-like LLTP. Compared to Vtg-like LLTP, however, only few additional encoded apoB-like LLTP have been found as yet in a single species, *e.g.* ARP in insects, an additional apoB in zebrafish (acc. CAH68927.1; ~50% similarity to acc. XP_694827), and a second Vtg gene in the decapod crustacean *Metapenaeus ensis* (Tsang *et al.*, 2003; Kung *et al.*, 2004). No such putative duplicates were observed for MTP. This indicates that

Table 2. Identified LLTP and their functions.

Subfamily	LLTP	Function	References
MTP	MTP	- facilitate biosynthesis of other LLTP	- Chapter 3; Shibata *et al.*, 2003; Sellers *et al.*, 2005; Shelness and Ledford, 2005
		- lipid antigen presentation (mammals)	- Brozovic *et al.*, 2004; Dougan *et al.*, 2005
apoB-like LLTP	apoB	- lipid transport to maintain homeostasis	- Frayn, 2003
		- vitellogenesis / placenta	- Terasawa *et al.*, 1999; Walzem *et al.*, 1999; Madsen *et al.*, 2004
		- vision: place lipids in Bruch's membrane?	- Li *et al.*, 2005a, b
	apoLp-II/I	- lipid transport to maintain homeostasis	- Van der Horst *et al.*, 2001, 2002; Van der Horst and Ryan, 2004
		- vitellogenesis	- Kawooya and Law, 1988; Kawooya *et al.*, 1988; Sun *et al.*, 2000
		- coagulogen	- Duvic and Brehélin, 1998
		- immunity: β-1,3-glucan recognition	- Duvic and Brehélin, 1998
		- immunity: lipopolysaccharide detoxification	- Ma *et al.*, 2005
		- morphogen transport during embryogenesis	- Panakova *et al.*, 2005
	ARP	?	
	M-177	?	- Epton *et al.*, 1999
	decapodan Vtg	- vitellogenesis	- Tsutsui *et al.*, 2000, 2004; Abdu *et al.*, 2002
Vtg-like LLTP	Vtg	- vitellogenesis	- Sappington and Raikhel, 1998a; Matsumoto *et al.*, 2003; Kato *et al.*, 2004; Romano *et al.*, 2004
	MEP	- immune system: regulation melanogenesis	- Lee *et al.*, 2000
	CP	- coagulogen	- Hall *et al.*, 1999; Yeh *et al.*, 1999

duplicates from the apoB-like and MTP subfamilies less often than Vtg-like LLTP attain a selective advantage.

To obtain a more clear view on the role of neofunctionalization in the establishment of LLTP diversity, the functions ascribed to the distinct LLTP were compiled (Table 2). The overview suggests several instances of neofunctionalization. For example, MEP is an insect Vtg-like LLTP that appears to have a function in regulating melanogenesis reactions in insect immunity (Lee *et al.*, 2000), unlike any other LLTP. In view of the presence of another Vtg-like LLTP in insects, *i.e.* Vtg, MEP appears to originate from a duplicated Vtg gene that has been retained due to its neofunctionalization. The single Vtg-like LLTP identified in decapodan Crustacea to date,

CP, may represent another example of neofunctionalization, as this Vtg-like LLTP functions as the major hemolymph coagulogen rather than as a yolk protein (Hall *et al.*, 1999; Yeh *et al.*, 1999; Tsutsui *et al.*, 2000, 2004; Abdu *et al.*, 2002). In the same taxon, an apoB-like LLTP, named Vtg, constitutes the major yolk protein, rather than a Vtg-like LLTP as in most animals. This likely reflects a shift in the relative importance of yolk proteins, as apoB-like LLTP have also been reported to engage in nutrient supply, especially of lipids, to the developing oocyte in insects as well as non-mammalian vertebrates (Kawooya and Law, 1988; Kawooya *et al.*, 1988; Walzem *et al.*, 1999; Sun *et al.*, 2000). Rather than a duplicate, decapod CP may as well be the Vtg ortholog that has been retained by neofunctionalization in the face of redundancy, as an apoB-like LLTP became the major yolk protein. Opposite to decapodan Crustacea, however, most insects depend on an apoB-like LLTP (apoLp-II/I) for coagulation and a Vtg-like LLTP (Vtg) as the major yolk protein (Chapter 2; Sappington and Raikhel, 1998a). Thus, members from phylogenetically distinct LLTP subfamilies have attained similar functions in vitellogenesis or coagulation in distinct taxa. This may reflect the taxon-specific division of multiple functions that used to be performed by the common ancestor of current apoB- and Vtg-like LLTP, although it is well possible that these functional similarities arose independently and are the result of parallel adaptations.

Conclusions and future directions

Examination of available sequence resources revealed two novel LLTP in insects, *i.e.* MEP and ARP. Phylogenetic analysis of selected LLTP, using conserved segments in the LLT module, revealed three major subfamilies of LLTP, *i.e.* MTP, Vtg-like LLTP, and apoB-like LLTP. Strikingly, the major yolk protein in decapodan Crustacea, which was also named vitellogenin, is not a Vtg-like but an apoB-like LLTP, exemplifying the functional flexibility within the family of LLTP. Despite the considerable amount of LLTP sequences that have become available since the recognition of this family (Babin *et al.*, 1999), the early origin and evolution of this family and its specific structural elements, *i.e.* the LLT module, vWF-D module and amphipathic clusters, cannot be established yet. Current efforts on additional metazoan genomes, particularly from the ancient metazoan phyla Cnidaria and Porifera, may provide further insight into the early origin and diversification of LLTP. Moreover, a wealth of LLTP functionality may still await discovery, as illustrated by the very recent discoveries that apoLp-II/I lipoprotein transports morphogen proteins during early embryonal development of *D. melanogaster* (Panakova *et al.*, 2005), and that mammalian MTP facilitates lipid antigen presentation (Brozovic *et al.*, 2004; Dougan *et al.*, 2005). For example, it remains to be established whether the adaptive advantage of multiple Vtg genes relates to new functions for Vtg or only to increased Vtg production (Bownes *et al.*, 1991), whether apoLp-II/I expression fulfils a function in the insect eye (Bogerd *et al.*, 2000), and what function(s) the vWF-D

module and pfam06448 motif have. Moreover, the function of the newly established insect LLTP, ARP, remains to be discovered. In particular, the functional implications of the observed structural variations in diverse LLTP remain to be established in most cases. Accordingly, the LLTP family provides a rich substratum for future integrative and comparative studies on protein evolution and adaptation throughout the animal kingdom.

Acknowledgements

The authors thank Jere P. Segrest and Martin K. Jones (Department of Medicine, University of Alabama at Birmingham Medical Center) for providing the program LOCATE.

Chapter 5

Sequence analysis of the non-recurring C-terminal domains shows that insect lipoprotein receptors constitute a distinct group of LDL receptor family members

Kees W. Rodenburg, Marcel M.W. Smolenaars, Dennis van Hoof, and Dick J. van der Horst

Biochemical Physiology, Department of Biology and Institute of Biomembranes, Utrecht University, The Netherlands

based on *Insect Biochem. Mol. Biol.* (2006) **36,** *in press*

Abstract

Lipoprotein-mediated delivery of lipids in mammals involves endocytic receptors of the low-density lipoprotein (LDL) receptor (LDLR) family. In contrast, insects were thought not to use endocytic receptors in lipid transport, as their lipoprotein, lipophorin, functions as a reusable lipid shuttle in lipid delivery. However, recent data indicate additional endocytic uptake of lipophorin, mediated by a lipophorin receptor (LpR) of the LDLR family. The two N-terminal domains of LDLR family members are involved in ligand binding and dissociation, respectively, and are composed of a mosaic of multiple repeats. The three C-terminal domains, viz. the optional O-linked glycosylation domain, the transmembrane domain, and the intracellular domain, are of a non-repetitive sequence. The present classification of newly discovered LDLR family members, including the LpRs, bears no relevance to physiological function. Therefore, as a novel approach, the C-terminal domains of LDLR family members across the animal kingdom were used to perform a sequence comparison analysis. The LpRs appeared to segregate into a specific group distinct from the groups encompassing the other family members, as each of the three C-terminal domains of the insect receptors is composed of a unique set of sequence motifs. Based on the present sequence motifs and their organization in the domains, LpR resembles most the groups of the LDLRs, very low-density lipoprotein (VLDL) receptors, and vitellogenin receptors. However, in sequence aspects in which LpR deviates from these three receptor groups, it most notably resembles LDLR-related protein-2, or megalin. These features might explain the functional differences disclosed between insect and mammalian lipoprotein receptors.

Introduction

Lipoprotein-mediated transport of lipid in the circulation is a general mechanism in animals, where lipid is being used for storage or metabolic purposes. The lipoprotein in insects, lipophorin (Lp), is proposed to function as a reusable shuttle, in which Lp selectively unloads lipids at the cell surface, subsequently followed by a new round of lipid loading. Consequently, in this process, Lp is not taken up into the cells by endocytosis (see Van der Horst, 1990; Ryan and Van der Horst, 2000; Van der Horst *et al.*, 2002; Rodenburg and Van der Horst, 2005). On the other hand, lipid delivery in mammals finally results in the endocytic uptake of remnant lipoprotein by cells, which is mediated by members of the low-density lipoprotein (LDL) receptor (LDLR) family. After endocytic uptake, the complex of lipoprotein and receptor dissociates and the lipoprotein undergoes lysosomal degradation, whereas the receptor is recycled back to the cell surface for another round of lipoprotein uptake (see Goldstein *et al.*, 1985; Krieger and Herz, 1994; Herz and Bock, 2002; Jeon and Blacklow, 2005). In view of the reusable

shuttle functioning of Lp, there would be no need for insects to express LDLR family members that can endocytose lipoproteins. However, recently it has become apparent that in insects also endocytic uptake of Lp is occurring, which is mediated by Lp receptors (LpRs) (Dantuma *et al.*, 1999; Lee *et al.*, 2003; Seo *et al.*, 2003; Van Hoof *et al.*, 2003, 2005a) (see Van der Horst *et al.*, 2001; Van der Horst *et al.*, 2002; Van der Horst and Ryan, 2004; Rodenburg and Van der Horst, 2005). Remarkably, sequence comparison indicates that these LpRs belong to the LDLR family (Dantuma *et al.*, 1999; Lee *et al.*, 2003; Seo *et al.*, 2003; Van Hoof *et al.*, 2003, 2005a, b). In addition, not only the LDLR family members of insects and mammals are homologous, but also their lipoprotein ligands (Babin *et al.*, 1999; Chapter 4). Strikingly, LpR of *Locusta migratoria* (Lm LpR), the only LpR for which the endocytic process to date has been characterized in details at the cellular level, appears to be able to recycle its ligand, a function not reported for any other LDLR family member (Van Hoof *et al.*, 2002, 2003, 2005a, b; see Van der Horst *et al.*, 2002; Van der Horst and Ryan, 2004; Rodenburg and Van der Horst, 2005).

LpRs are LDLR family members and are composed of the five characteristic domains (see Figure 7 in Chapter 1), also present in human LDLR: (1) an N-terminal ligand-binding domain, (2) an epidermal growth factor (EGF)-precursor homology domain, (3) an optional O-linked glycosylation domain (OLGD), (4) a single transmembrane spanning segment or transmembrane domain (TMD), and (5) a C-terminal intracellular domain (ICD). The ligand-binding domain is composed of multiple cysteine-rich repeats that are involved in binding of different ligands, including lipoproteins. The EGF-precursor homology domain is composed of different repeats that function in ligand dissociation after endocytic uptake of the ligand. The OLGD, present in some but not all family members, contains several serine and threonine residues that potentially are O-linked glycosylated. The TMD spans the cell membrane once, to position the receptor properly during intracellular transport and at the cell surface. The ICD mediates the endocytic uptake of receptor in complex with lipoprotein via the NPXY sequence motif in the latter domain that acts as a clathrin-coated pit internalization signal (see Willnow, 1999; Herz and Bock, 2002; Schneider and Nimpf, 2003). Because they are composed of non-recurring sequences, unlike the ligand binding and EGF-precursor homology domains, we refer to the OLGD, TMD and ICD as the non-recurring C-terminal domains.

The family of LDLR members is composed of a large variety of receptors that, most notably, contain a different number of cysteine-rich ligand-binding repeats in the first domain, often in combination with multiple copies of the second (EGF-precursor homology) domain, thereby allowing the binding and dissociation of different ligands, both in number and specificity (Herz and Bock, 2002). In addition to their endocytic role, several family members have been shown to be involved in other processes: they effect signal transduction (LDLR-related protein, LRP; very low-density lipoprotein receptor, VLDLR; and LRP8 [apolipoprotein E receptor-2, apoER2]), mediate polarized expression in cells (LDLR; LRP2, also known as megalin), perform transcytosis (LRP2), ligand

recycling (LpR), activate transcription (LRP; LRP2), or interact with intracellular proteins for which a cellular event has not been identified yet (Herz and Strickland, 2001; Herz and Bock, 2002; May and Herz, 2003; Schneider and Nimpf, 2003; Strickland and Ranganathan, 2003; Van der Horst and Ryan, 2004; Rodenburg and Van der Horst, 2005). Involvement in function has been characterized for the OLGD of LRP8 (May *et al.*, 2002, 2003), and the ICD of LDLR, VLDLR, LRP, LRP2, and LRP8 (Matter *et al.*, 1992, 1994; Trommsdorff *et al.*, 1999; Djordjevic *et al.*, 2000; Gotthardt *et al.*, 2000; Li *et al.*, 2000; Oleinikov *et al.*, 2000; Rader *et al.*, 2000; Stockinger *et al.*, 2000, 2002; Patrie *et al.*, 2001; He *et al.*, 2002; Melman *et al.*, 2002; Mishra *et al.*, 2002; Larsson *et al.*, 2003; Nagai *et al.*, 2003; Petersen *et al.*, 2003; Takeda *et al.*, 2003). Tissue specific splice variants in the OLGD have been identified for VLDLR, LRP8 (Martensen *et al.*, 1997; Iijima *et al.*, 1998; Nakamura *et al.*, 1998; Clatworthy *et al.*, 1999; Magrane *et al.*, 1999; Sun and Soutar, 1999; Korschineck *et al.*, 2001), and LpR from the insects *Aedes aegypti* (Seo *et al.*, 2003) and *Galleria mellonella* (Lee *et al.*, 2003), suggesting alternative functioning of these variants. Thus, it appears that each LDLR family member performs its own specific set of functions, and that the OLGD and ICD of individual receptors are directly involved in the execution of many of these functions (see Willnow, 1999; Herz and Strickland, 2001; Herz and Bock, 2002; Andersen and Petersen, 2003; May *et al.*, 2003; Schneider and Nimpf, 2003; Stolt *et al.*, 2003; Strickland and Ranganathan, 2003; Jeon and Blacklow, 2005).

Mammalian LDLR family members are often compared and classified based on the number of the N-terminal ligand-binding domains and EGF-precursor homology domains, and the number of repeats within these domains (see Willnow, 1999; Herz and Bock, 2002; Jeon and Blacklow, 2005). For example, the single ligand-binding domain of mammalian LDLR and VLDLR contain 7 and 8 N-terminal cysteine-rich repeats, respectively. Accordingly, Lm LpR has been classified as belonging to the group of VLDLR due to its eight cysteine-rich ligand-binding repeats (Dantuma *et al.*, 1999). However, the functional and evolutionary significance of such a classification is lacking, as the number of repeats is known to change with relatively high frequency (Sappington and Raikhel, 1998b) and the number of ligand-binding repeats in *A. aegypti* LpR (Aa LpR) can be modified by splicing (Seo *et al.*, 2003). Thus, comparison of repeat numbers in the ligand-binding and EGF-precursor homology domains does not suffice for the classification of non-mammalian LDLR family members, such as insect LpRs. A more profound categorization would be based on sequence homology. Unfortunately, the assignment of homologous repeats is particularly difficult due to rearrangements, duplications and deletions in the repeat regions during evolution (Sappington and Raikhel, 1998b)

The non-recurring domains, *i.e.* the OLGD, TMD and ICD, may provide another tool for the classification of LpR among other LDLR family members. Therefore, a sequence analysis was performed on the non-recurring domains in LpR, as compared to

other LDLR family members. The observed sequence motifs are discussed in the context of their meaning for the functioning of the LpR group members.

Methods

Amino acid sequences of endocytic LDLR family members that contain a TMD, and at least one clathrin-coated pit internalization signal, [F/Y]XNPXY (amino acids are represented by the one letter code; X stands for any amino acid; letters enclosed in square brackets indicate a position were either of the represented amino acids is allowed) were obtained with NCBI Entrez Protein search engine (http://www.ncbi.nlm.nih.gov/entrez/). LpRs were chosen that either were demonstrated to bind lipophorin or are expressed in fat body tissue. The two putative LpRs from *Drosophila melanogaster* (Dm) and *Anopheles gambiae* (Ag) (Table 1) were included after an analysis with the LpRs from *L. migratoria* (Lm), *G. mellonella* (Gm), and *A. aegypti* (Aa), since their sequence motifs comply with all aspects characterized in this study. Amino acid sequence alignment was based on the region starting immediately after the sixth cysteine (C) residue of EGF-precursor homology repeat C to the very C-terminus. In this study the first residue of this region is indicated as residue 1.

The boundaries of the TMD of the receptors were determined individually using the on-line transmembrane regions prediction software TMpred at ch.EMBNET.org (http://www.ch.embnet.org/software/TMPRED_form.html) (Hofmann and Stoffel, 1993). The sequences were analyzed for potential glycosylation sites in the OLGD, and casein kinase-2 (CK2) phosphorylation, cAMP/cGMP dependent protein kinase (cA/GDPK) and PKC-phosphorylation sites in the ICD, using the Prosite scan software (Sigrist *et al.*, 2002) at the Expasy proteomics server.

The multiple amino acid sequence alignment of the domain regions was created with the on-line CMBI Clustal-W server (http://www.cmbi.kun.nl/bioinf/tools/clustalw.shtml), and manually modified by aligning potential N-linked glycosylation sites in the OLGD or ectodomain segment, the proline, glycine and aromatic residues in the TMD, and the sequence motifs I to VII in the ICD (Figure 1). The modified alignment was used to reconstruct an unrooted tree by quartet puzzling and maximum likelihood (Strimmer and von Haeseler, 1996), based on 10,000 puzzling steps with the on-line software Puzzle (http://bioweb.pasteur.fr/seqanal/interfaces/Puzzle.html), the infile data tree of which was visualized with TreeView 1.6.6 (http://www.taxonomy.zoology.gla.ac.uk/rod/treeview.html) (Page, 1996). The branch lengths are proportional to the degree of relative sequence divergence.

Table 1. Reference information of vertebrate and invertebrate LDLR family members. The accession numbers refer to database information linked to Entrez-Pubmed.

Species	Receptor	Abbreviation	Accession
Homo sapiens (human)	LDLR	Hs LDLR	NP_000518
	VLDLR	Hs VLDLR	NP_003374
	LRP	Hs LRP	NP_002323
	LRP2	Hs LRP2	NP_004516
	LRP8	Hs LRP8	NP_004622
Mus musculus (mouse)	LDLR	Mm LDLR	QRMSLD
	VLDLR	Mm VLDLR	P98156
	LRP	Mm LRP	NP_032538
	LRP2	Mm LRP2	XP_130308
	LRP8	Mm LRP8	NP_444303
Rattus norvegicus (Norway rat)	LDLR	Rn LDLR	NP_786938
	VLDLR	Rn VLDLR	NP_037287
	LRP2	Rn LRP2	NP_110454
Oryctolagus cuniculus (rabbit)	LDLR	Oc LDLR	P20063
	VLDLR	Oc VLDLR	P35953
Sus scrofa (pig)	LDLR	Ss LDLR	AAC39254
Cricetulus griseus (Chinese hamster)	LDLR	Cg LDLR	P35950
Bos taurus (cow)	VLDLR	Bt VLDLR	NP_776914
Gallus gallus (chicken)	LRP	Gg LRP	P98157
	LR8B	Gg LR8B	CAA65729
	VGR	Gg VGR	P98165
Danio rerio (zebrafish)	LDLR	Dr LDLR	AAP22970
Chiloscyllium plagiosum (bambooshark)	LDLR	Cp LDLR	AAB42184
Oreochromis aureus (blue tilapia)	VGR	Oa VGR	AAO27569
Oncorhynchus mykiss (rainbow trout)	VGR	Om VGR	CAD10640
	SLR	Om SLR	CAA05874
Xenopus laevis (African clawed toad)	LDLR	Xl LDLR	Q99087
	VLDLR	Xl VLDLR	JC4858
Drosophila melanogaster (fruit fly)	GH26833	Dm LpR	AAQ22563
Locusta migratoria (migratory locust)	LpR	Lm LpR	CAA03855
Galleria mellonella (wax moth)	LpR	Gm LpRov	n/a
Aedes aegypti (yellow fever mosquito)	LpR	Aa LpRov	AAK72954
Anopheles gambiae (malaria mosquito)	LpR	Ag LpR	EAA03681
Caenorhabditis elegans (nematode)	LRP-1	Ce LRP1	NP_492127
	RME-2	Ce RME2	NP_500815

Results and Discussion

Amino acid sequence alignment of the non-recurring C-terminal domains of LDLR family members

To disclose common and unique sequences and motifs in the C-terminal domains, selected LDLR family members (Table 1) were aligned (Figure 1). The sequences of the TMD and OLGD (absent in LRP, LRP2, and both *C. elegans* (Ce) receptors) appeared to be highly specific for each group, as illustrated by the recognition of intergroup sequence similarity only in the TMD between VLDLR and VGR family members. In contrast, the ICD also contained muliple sequence motifs that were present in most LDLR family members, in addition to group-specific sequence motifs. Considering their small size and possible functional value, the common motifs might have arisen independently among different LDLR groups. Thus, the non-recurring C-terminal domains of LDLR family members may not be suitable for establishing the evolutionary relationships among different LDLR family groups.

Tree construction was applied to visualize the diversity and similarity among LDLR family members in the non-recurring domains, which can readily be observed in the alignment (Figure 1). Strikingly, the topology of the obtained tree (Figure 2) reflects the generally accepted grouping of vertebrate LDLR family members that have been subdivided according to number of cysteine-rich repeats and domain arrangement. In addition, VGRs and VLDLRs clustered into a clade (78% support), in accordance with a previous phylogenetic analysis of cysteine-rich repeats in the ligand-binding domain (Sappington and Raikhel, 1998b). The long branches of both *C. elegans* LDLR members concur with their lack of similarity to any other included sequence. Strikingly, LpRs form an isolated branch that diverges from other lipoprotein receptors (80% support). Based on their eight cysteine-rich ligand-binding repeats (Dantuma *et al.*, 1999; Cheon *et al.*, 2001), Lm LpR and Aa LpRov would be classified as belonging to the group of VLDLR. In contrast, the present sequence analysis of their non-recurring C-terminal domains demonstrates that LpRs constitute a unique group of LDLR family members.

Since the non-recurring C-terminal domains have been reported to engage in the execution of many of the functions of multiple LDLR family members (see Willnow, 1999; Herz and Strickland, 2001; Herz and Bock, 2002; Andersen and Petersen, 2003; May *et al.*, 2003; Schneider and Nimpf, 2003; Stolt *et al.*, 2003; Strickland and Ranganathan, 2003; Jeon and Blacklow, 2005), analysis of the unique set of C-terminal sequences in LpRs may help to explain their functioning. In the subsequent paragraphs, the sequence motifs and their functional implications are discussed for each non-recurring domain, in the context of the meaning for the functioning of the LpR group members.

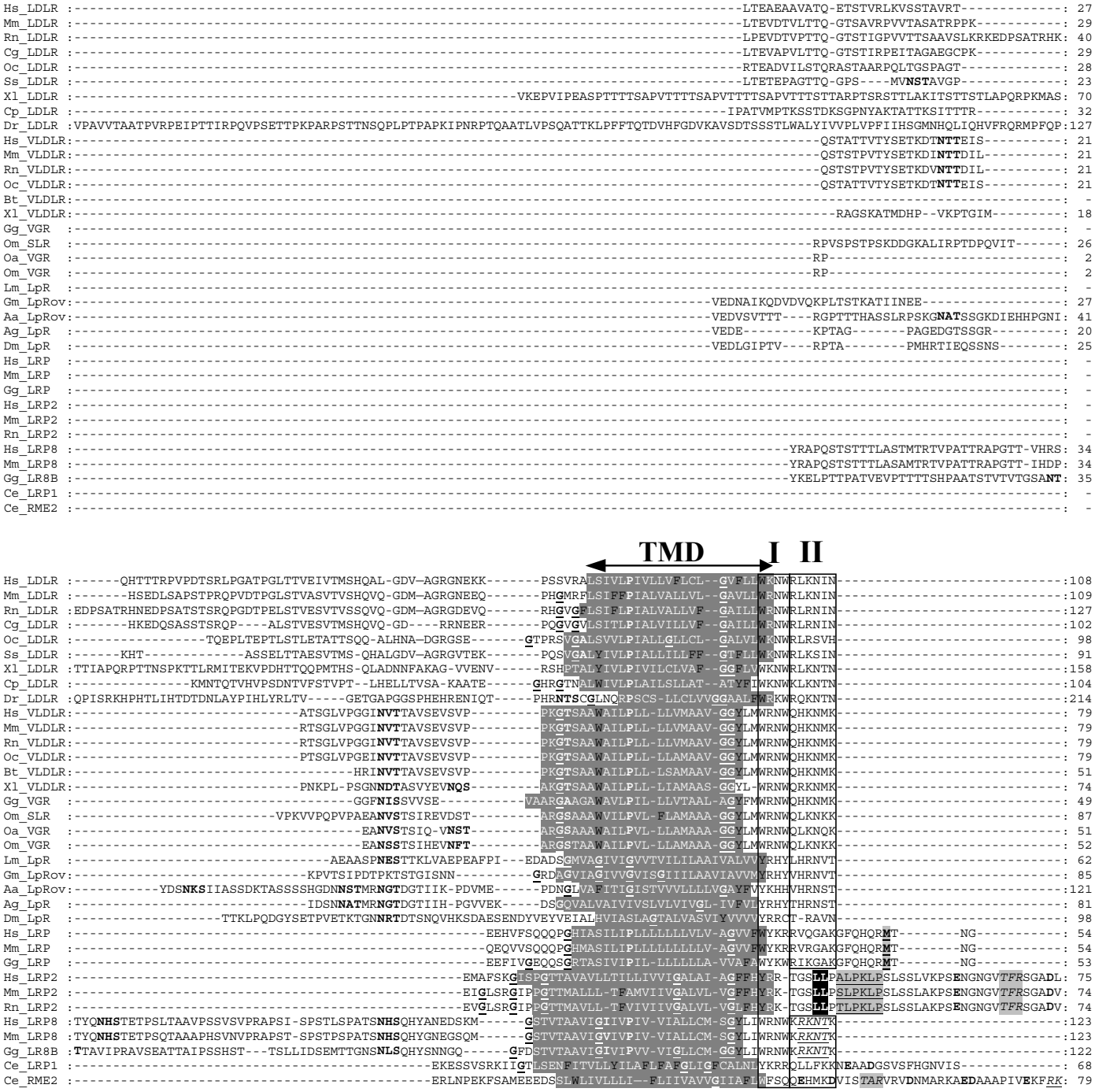


Figure 1. Alignment of sequence motifs in the non-recurring C-terminal domains of LDLR family members. For each individual receptor, the sequence C-terminal to the last cysteine residue of the most C-terminal EGF-type repeat is included. In the alignment, the predicted TMD of each LDLR is indicated by a dark grey background. The sequences preceding the TMD correspond to the OLGD or, for LRP, LRP2 and *C. elegans* (Ce) family members, sequences flanking the most C-terminal EGF-type repeat. The sequences after the TMD, at its right side, correspond to the ICD. Where possible, sequences of the same LDLR family groups were aligned, even if alignment among all groups was impossible. Family members are identified on the left of each line of sequence, according to their abbreviations in Table 1. Residue numbering is indicated on the →

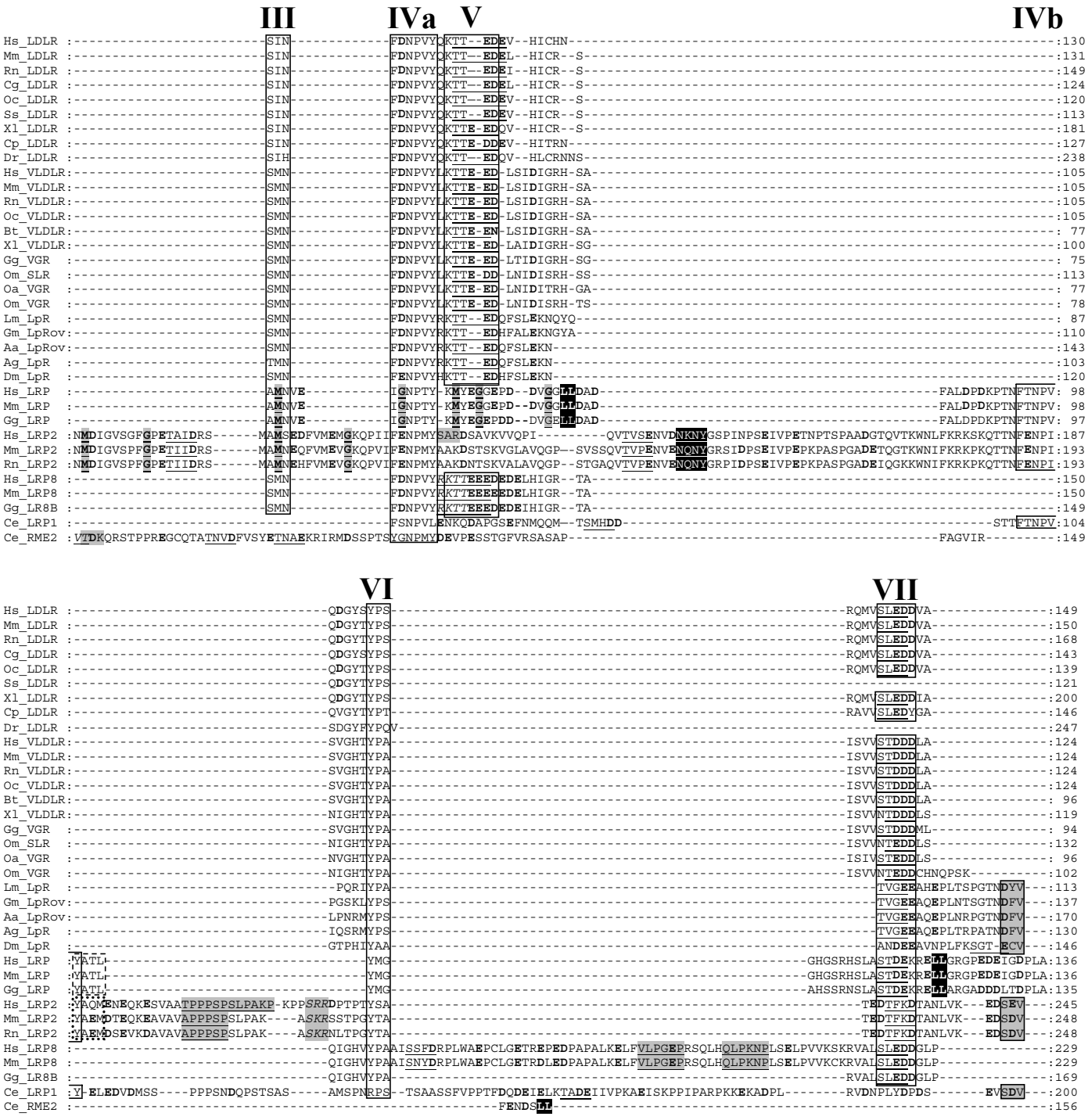


Figure 1. (continued from previous page)

right of each line. In the OLGD, sequence motifs for N-linked glycosylation (NX[S/T]) are shown in bold face. In the TMD, P residues are indicated in white bold face, aromatic residues in black, and G residues in bold face with underlining (also in the proximate sequences N-terminal of the TMD). In the ICD, roman numerals above the alignment mark seven motifs (I to VII) that are present in most members. In the alignment, these common motifs have been indicated by boxes around the involved sequences. Motif I overlaps with the predicted TMD for some LDLR family members. Motif IV, the motif for clathrin-coated pit internalization, is present at two distinct locations, marked as IVa and IVb. Motif IVb is present only in the LRP and LRP2 receptors, and Ce LRP1. The acidic residues (E and D) in the ICD are in bold face, even if present in motifs with other →

The O-linked glycosylation domain (OLGD)

The OLGD is located between the EGF-precursor homology domain and the transmembrane domain, and is present in all compared LDLR family members, except in LRP, LRP2, and both Ce receptors, where it is rudimentary or absent. The OLGD contains multiple serine (S) and threonine (T) residues, several of which are, putatively, O-linked glycosylated (*e.g.* Willnow, 1999; Herz and Bock, 2002). This domain does not contain sequence motifs that are present in all family members that contain such a domain. Possibly, an exception forms a glycine (G) residue that is present at the boundary of the OLGD and TMD in almost all members, including the LRPs (Figure 1). In general, G may introduce flexibility in a protein structure, which could be in line with a structural transition between the OLGD or the ectodomain, and the TMD. Within the different receptor groups, sequence identity is observed; however, there is substantial sequence length variation among the members of a group. A sequence motif for the N-linked glycosylation of asparagine (N) residues, *i.e.* NX[S/T], is present in the groups for LpR, VLDLR, VGR, LRP8, and two members of the LDLR group (Figure 1; characters in bold face). This motif occurs one to four times per receptor. The members of the LpR group, clearly also harboring an OLGD, show the largest variation in occurrence number for this motif. Lm LpR and *G. mellonella* (Gm) ovary LpR (Gm LpRov) contain one such motif, whereas *A. aegypti* fat body LpR (Aa LpRfb) contains four. Among the different groups, one such motif seems to be positioned within a fixed distance from the transmembrane domain (about 10 residues). Potentially, this site becomes glycosylated, since a minimal distance of about 10 amino acids is required between the membrane boundary and the N residue, for it to become glycosylated (Monné and von Heijne, 2001). The site in LDLR from pig (Ss LDLR), and VLDLR from a toad (Xl VLDLR), and VGR from tilapia and trout (Oa VGR and Om VGR, respectively) are located closer to the membrane. Therefore, they are most likely not glycosylated. A function for the N-linked glycosylation in this domain is currently not available. However, since N- and O-linked glycosyl chains differ dramatically in carbohydrate composition, the N-linked sugar may indirectly alter or introduce binding potential of the receptor. It was long thought that the

markings. Additional motifs in ICD sequences are marked by a unique combination of face (italics, bold, or underlining), background color, or surrounding lines. Underlining only: CK2 phosphorylation motif ([S/T](X2)[D/E]); underlined and in italics: cA/GDPK motif ([K/R][K/R]X[S/T]); in italics on a grey background: PKC dependent phosphorylation motif ([S/T]X[K/R]); bold and underlined M and G residues on a grey background: M(X9)G motifs in LRP and LRP2; underlined on a grey background: potential SH3 binding sites (XØPXXP) in the LRP2 and LRP8 receptors; grey background and surrounded by black solid lines: putative PDZ-domain interaction sequences in the LpR and LRP2 receptors, and Ce LRP1; white bold face characters on a black background: dileucine (LL) motifs in LRP, LRP2, and Ce RME2; white underlined characters on a black background: potential NPXY motif (NXXY) in the LRP2 receptors; surrounded by dashed lines: YXXØ motif in LRP receptors; surrounded by dotted lines: SH2 binding motif (YXXM) in LRP2 receptors. The Y-residue of the latter two motifs also participates in the preceding motif IVb. The character Ø in above motifs represents a residue with a bulky hydrophobic side chain.

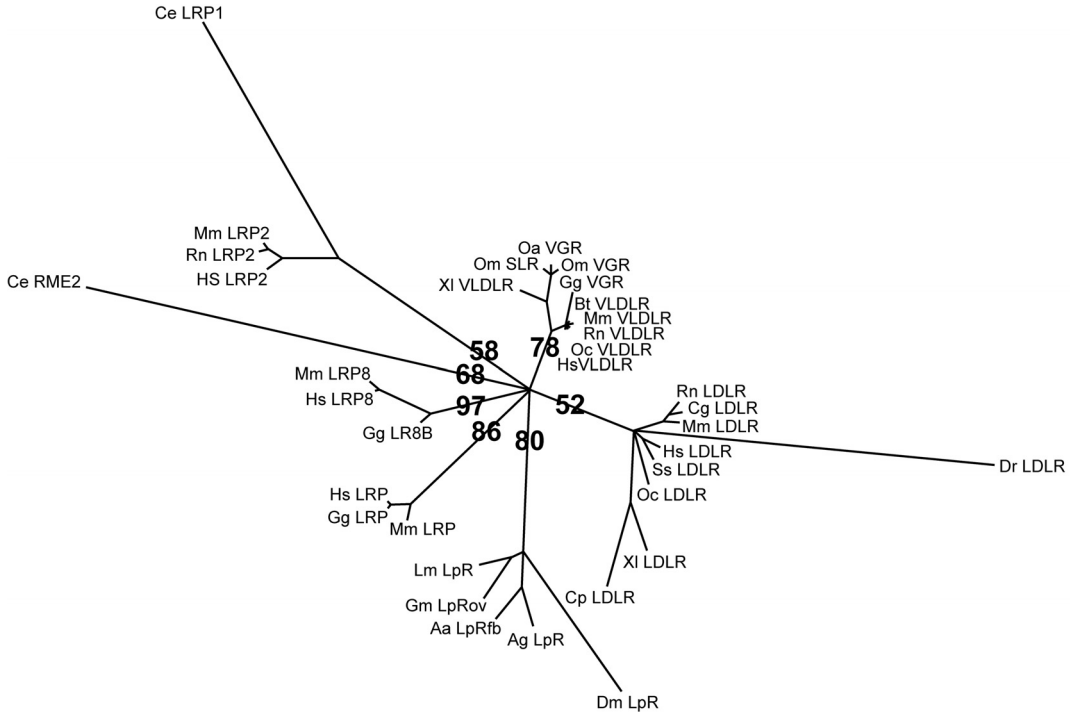


Figure 2. Tree representation of the diversity in the non-recurring domains in members of the endocytic LDLR family. Numerals in the tree indicate the support, expressed as a percentage, for the shown grouping. Branch lengths are proportional to the degree of relative sequence divergence. See Table 1 for abbreviations of lipoprotein receptors.

domain would serve a mere structural role in spacing the ectodomain with the transmembrane domain (Herz and Bock, 2002), and potentially promoting stability of the receptor (Kingsley *et al.*, 1986; Magrane *et al.*, 1999). However, tissue specific splice variants in this domain of human VLDLR and LRP8 (Martensen *et al.*, 1997; Iijima *et al.*, 1998; Nakamura *et al.*, 1998; Clatworthy *et al.*, 1999; Magrane *et al.*, 1999; Sun and Soutar, 1999; Korschineck *et al.*, 2001), and Gm LpR (Lee *et al.*, 2003) and Aa LpR (Seo *et al.*, 2003), point to a specific functional role for this domain. In addition, recent data for LRP and LRP8 suggest a role for this domain in the regulation of the release of the intracellular domain from the plasma membrane, and thus the regulation of signaling pathways that depend on the release of the cytoplasmic domain from the plasma membrane (May *et al.*, 2002, 2003). VLDLRs and LpRs, therefore, may also be involved in such signaling events.

The transmembrane domain (TMD)

The TMD plays an important role in receptor functioning, since it positions the receptor in the membrane during different intracellular transport routes (*e.g.* ER and Golgi transport), and numerous recycling passes between plasma membrane and various endocytic vesicle systems (see Maxfield and McGraw, 2004). All these membrane systems may have their

specific requirements for the receptor protein-lipid interaction. The TMD seems crucial for the correct positioning and subsequent functioning of its flanking domains, *i.e.* the ICD and OLGD. It putatively forms a transmembrane α-helix, and functions as a membrane anchor (Krieger and Herz, 1994; Willnow, 1999; Herz and Bock, 2002). The sequence of this domain is relatively enriched in the hydrophobic amino acid residues alanine (A), valine (V), isoleucine (I), leucine (L), aromatic residues (phenylalanine [F], tyrosine [Y], or tryptophane [W]), and G, and varies in length between 19 and 32 residues (Figure 1). Both the N- and C-terminal boundary of the TMD seems to be marked by a G residue. Notably, this G residue is not always predicted to be part of the TMD. In several cases, the C-terminal boundary is determined by a W or Y residue, or by a residue that is one to three positions away from this aromatic residue. The sequence of the TMD is markedly identical among the individual members of a group, but not at all among the different receptor groups.

Five noteworthy features can be discerned in the sequence comparison of this domain. First, close to the N-terminal boundary of the groups of LDLR, VLDLR, VGR, and both Ce receptors an aromatic residue is found. These residues, and W in particular, are known to anchor transmembrane helices to the interface of phospholipid bilayer and extracellular space (Hessa *et al.*, 2003; see Killian and von Heijne, 2000; Killian, 2003). For the LRP, LRP2, LRP8, and LpR groups, such a residue is absent, and the helix may therefore be more flexible in movement on the N-terminal side. A second feature is the presence of P in the N-terminal part of the domain of all groups, except that of the insect receptors, LPR2 and both Ce receptors. P is known to act as a helix breaker; however, it is tolerated at the start of an α-helix. The disturbance of the helix structure induced by P may be crucial to the function of the domain. Three of the five LpRs seem to compensate for the lack of P by having (spaced) G residues. G is also known to be a helix breaking residue. Notably, in Lm LpR and Gm LpRov the G-residues are spaced three residues from each other (GXXXG motif). This motif is unique to the insect group, and its role is currently unknown. However, such a motif has been observed in the TMD of other receptors, where the position of the motif at the N- or C-terminal boundary of the domain causes the ErbB1 and 2 receptors to hetero- or homodimerize, respectively (Gerber *et al.*, 2004). In this respect it is important to note that several LDLR family members have been shown to function in signal transduction, and interact with receptors that are unrelated to the LDLR family (Herz and Bock, 2002; Schneider and Nimpf, 2003). However, dimerization of any of the family members has not been reported, to date. Third, in all groups the middle part of the TMD is characterized by many large aliphatic residues (V, L, and I) that are excellently fit to form the core of the α-helical structure, crossing the phospholipid bilayer. The fourth feature is the presence of one or two G residues, in combination with one or more aromatic residues, in the C-terminal TMD part in members of all groups. The insect LpR group members deviate most among each other at this point, lacking one of the two, or both residues. The role of the combination of the two residues

may be introducing flexibility in the helix, to allow movement of the TMD. Last, the C-terminal boundary is marked in a one-to-three residue distance from the ubiquitous first aromatic residue in the recognized motif I (Figure 1). In some receptors, this includes the positively charged, basic, residue of motif II. We do not know whether this variation in C-terminal boundary reflects a functional difference or has to do with the formulation of the algorithm behind the transmembrane segment prediction software. In the receptors in which the aromatic residue of motif I is not part of the TMD, often a Y precedes this residue, and Y is then identified as the boundary of the domain. This is in good agreement with Y being able to substitute with W at a lipid-water interface position (Killian, 2003). Nevertheless, considering the observation that the combination of an aromatic residue with a positively charged residue (arginine [R] or lysine [K]) yields maximally stabilized membrane anchoring, due to the additional interaction of the positive charge with the negatively charged phospholipid head group (Hessa *et al.*, 2003; see Killian and von Heijne, 2000; Killian, 2003), it may well be possible that the first two residues of the tetrapeptide, [W/Y][K/R]X[W/Y], motif I, participate importantly in the functioning of the TMD. The movement of this domain could serve to position the intracellular domain for correct interaction with cell machinery, or adapt the receptor to bilayer thickness during the different transport routes.

The intracellular domain (ICD)

The ICD is exposed to the cytoplasm and plays an indispensable role in the endocytic process, via the clathrin-coated pit internalization motif, [F/Y]XNPXY (Figure 1; motif IVa). It functions in intracellular vesicle sorting and for several individual family members it has also been shown to function in signal transduction, either directly by interacting with specific cytoplasmic adaptor proteins, or indirectly by the interaction of the extracellular ligand-binding domain to ligands that function in signal transduction. Sequence motifs or interaction sites involved in these activities have been characterized in the ICD for individual receptors. Table 2 provides a summary of these motifs and their (potential) function. Together with figure 1, it provides an overview of the complexity and mosaic character of the ICDs for the different groups. Recently, several additional activities have been uncovered for individual receptors, which may point to the involvement of the ICD in even more interactions (*i.e.* processes): LRP (May *et al.*, 2002) and Lm LpR (Van Hoof *et al.*, 2002, 2003, 2005a, b).

Our sequence analysis shows that the ICD contains seven sequence motifs that are present in almost all groups (Figure 1). The ICDs of the “large” LDLR family members, *i.e.* LRP, LRP2, and Ce LRP1, are longer than those of the LDLR, VLDLR, VGR, and LpR groups, due to sequences adjacent to and interjacent of the seven motifs. The motifs IVa and b are motifs for clathrin-coated pit internalization. Only the LRP, LRP2, and both Ce receptors contain two of these motifs. The motifs V and VII are recognized as potential

CK2 phosphorylation sites. Phosphorylation may serve several functions, particularly in signal transduction; however, it has also been postulated to be involved in ubiquitination and subsequent degradation of cytoplasmic proteins (Djordjevic *et al.*, 2000). The specific role of the other four motifs (I, II, III, and VI) in receptor functioning is currently unknown. The interjacent sequences and the sequence C-terminally flanking motif VII are similar within each group. For several individual members a role has been identified for a number of sequence motifs (Table 2). The motifs are discussed only when relevant to LpR function.

With respect to the overall motif organization, LpRs resemble the family members of the LDLR, VLDLR, and VGR groups. The insect receptors deviate from these three groups at their N- and C-terminal boundaries, in which they resemble the receptors of the LRP2 group: 1) motif I begins with a Y-residue instead of W, and 2) the extreme C-terminal sequence contains a unique extension in addition to a PDZ-domain interacting sequence (Figure 1; Table 2).

Considering the overall distribution of sequence motifs in all the receptors, including the insect lipoprotein receptors, it appears that the intracellular domain is composed of an integrated network of motifs that overlap or intertwine, forming a mosaic of motifs. Noticeably, the potential CK2 phosphorylation site, motif V, overlaps with a sequence putatively involved in intracellular transport of human LDLR from ER to cell membrane, Q^{119}KTTEDE (Hoppe and Joiner, 2000) (Table 2). With respect to this motif, it is relevant to note that the first residue (*i.e.* Q for LDLR) is group specific, and replaced by R in the LpRs, introducing a cA/GDPK site in this group of receptors (Figure 1). A second example of such a mosaic of motifs of LDLRs is that the acidic patch (E^{145}DD) of the putative CK2 phosphorylation site of motif VII, together with G^{133} and Y^{134} N-terminal of motif VI, is involved in the basolateral sorting of the human receptor (Matter *et al.*, 1992, 1994). The insect receptors lack the GY dipeptide motif (Figure 1; Table 2). A third example is provided by the group of LRP receptors; three putative M(X9)G, an FDNPXY (motif IVa), an acidic patch, and a dileucine (LL) motif are found in close proximity to each other (Figure 1; Table 2). An M(X9)G motif, combined with the acidic patch and dileucine motif is involved in lysosomal degradation of the Batten disease protein CLN3 (Kyttala *et al.*, 2004). Also, in LRP2 receptors the M(X9)G motif and acidic patches are integrated with other motifs (Figure 1). Currently, only LRP is known to be targeted for proteasomal degradation (Melman *et al.*, 2002). The occurrence of two or more acidic patches seems to be another common feature of the family as a whole (Figure 1; all acidic residues in the ICD are indicated in bold face).

The role of such a mosaic kind of organization of the ICD may be to allow for regulation and control of (signaling) events by mutual exclusion of ICD binding proteins that direct mutually exclusive processes in which the receptor can be involved. The need for the mosaic kind of organization may be due to the relative short sequence length of the

Table 2. Overview of sequence motifs in the ICD of LDLR family members.

Receptor	Sequence/binding site [1)]	Interacting protein [2)]	Function [3, 4, 5)]
LDLR	- F^{112}DNPVY	- Dab1/Dab2/ARH/FE65	- CCPI (1, 2, 3)
(Hs)	- Y^{117}/E^{122}DE/E^{145}DD	- COPII proteins?	- Basolateral sorting (4, 5)
	- Q^{118}KTTEDE (motif V)		- Intracellular trafficking? (6)
	- $G^{133}Y^{134}$/E^{145}DD		- Basolateral sorting (4, 5)
	- [S/T]$^{121, 122, 143}$(X2)[D/E]	- CK2	- Potential phosphorylation of S or T
LpR	- F^{66}DNPVY		- CCPI?
(Lm)	- [S/T]$^{73, 95}$(X2)[D/E]	- CK2	- Potential phosphorylation of T
	- [K/R][K/R]X[S/T]^{72}	- cA/GDPK	- Potential phosphorylation of T
	- D^{111}YVc		- PDZ-domain interaction?
VLDLR	- F^{83}DNPVY	- Dab1	- CCPI (1)
(Hs)	- [S/T]$^{91, 92, 118, 119}$(X2)[D/E]	- CK2	- Potential phosphorylation of S or T
VGR	- F^{53}DNPVY (motif IVa)		- CCPI?
(Gg)	- [S/T]$^{61, 62, 88, 89}$(X2)[D/E]	- CK2	- Potential phosphorylation of S or T
LRP	- I^{60}GNPTY (motif IVa)		- CCPI? (1, 7)
(Hs)	- M^{51}(X9)G^{61}		- Degradation? (8)
	- M^{56}(X9)G^{70}		- Degradation? (8)
	- M^{67}(X9)G^{77}		- Degradation? (8)
	- DDVGGL^{79}LDAD	- Dab1/FE65	- Sorting?/Degradation? (8, 9)
	- F^{94}TNPVY (motif IVb)	- μ2 of AP-2	- CCPI (1, 7, 10)
	- Y^{99}XXØ [6)]		- CCPI (7, 10)
	- [S/T]^{115}(X2)[D/E]	- CK2	- Potential phosphorylation of S
	- L^{122}L	- β1 of AP-1?	- CCPI?/TGN sorting? (7, 10)
LPR2	- H^{37}YRR … IGV^{81} [7)]	- MegBP	- Vitamin D metabolism (11)
(Hs)		- Ankra	- Ankyrin binding (12)
	- L^{44}L [8)]		- Sorting? (8)
	- [S/T]$^{68, 117}$X[K/R]	- PKC kinase	- Potential phosphorylation of S or T
	- [S/T]$^{88, 132, 231}$(X2)[D/E]	- CK2	- Potential phosphorylation of S or T
	- M^{77}(X9)G^{85}		- Degradation? (8)
	- M^{96}(X9)G^{105}		- Degradation?
	- F^{111}ENPMY (motif IVa)	- Dab1/Dab2	- CCPI (1, 13)
	- VDN^{139}KNY	- Dab1?	- CCPI? (1)
	- F^{183}ENPIY (motif IVb)		
	- Y^{188}XXM		- SH2 domain interaction (14)
	- XØP$^{46, 204, 210}$XXP [6, 7)]		- SH3 domain interaction (14)
	- S^{243}Evc [9)]	- PSD95 [9)]	- PDZ-domain interaction (1, 15)
LPR8 [10)]	- [K/R][K/R]X[S/T]$^{119, 133}$	- cA/GDPK	- Potential phosphorylation of S or T
(Hs)	- F^{127}DNPVY	- Dab1	- CCPI (1)
	- [S/T]$^{135, 136, 161, 222}$(X2)[D/E]	- CK2	- Potential phosphorylation of S or T
		- JIP-1 and JIP-2	- Signal transduction (16)
	- A^{159}IS…KSK^{217} [11)]		- SH3 domain interaction (14)
	- XØP$^{193, 204}$XXP [6)]		
LPR1	- F^{69}SNPVL (motif IVa)		- CCPI?
(Ce)	- [S/T] ^{92}(X2)[D/E]	- CK2	- Potential phosphorylation of S
	- F^{100}TNPVY (motif IVb)		- CCPI?
RME2	- [S/T]$^{53, 81}$X[K/R]	- PKC kinase	- Potential phosphorylation of T
(Ce)	- [K/R][K/R]X[S/T]^{78}	- cA/GDPK	- Potential phosphorylation of T
	- [S/T]$^{97, 106}$(X2)[D/E]	- CK2	- Potential phosphorylation of T
	- Y^{121}GNPMY		- CCPI?
	- L^{155}L [12)]		- Sorting or PDZ-domain interaction

Table 2 *(Opposite page)***. Overview of sequence motifs in the ICD of LDLR family members.** Motifs, other than the motifs I to VI, except motif IVa and IVb, and their respective binding proteins (if known) are indicated. In all sections, a blank indicates that no information is available. Sequence motifs are indicated in italics if their function has been identified in a protein unrelated to the LDLR family. In this table, proteins are mentioned only if a binding site has been identified that is smaller then the complete sequence of the ICD. We refer to Gotthardt *et al.* (2000), and the reviews mentioned in the manuscript for most of the proteins not mentioned here. [1)] Amino acid sequence of the motif or residues. A number shown in superscript indicates the number of the residue in the alignment; multiple numbers indicate multiple sites. Motifs involving an X (any amino acid) are represented as consensus sequences, and the number(s) in superscript in the motifs mark the residue(s) used in Figure 1. [2)] Indicates the name of the protein that binds to the sequence motif. [3)] Indicates the function of the cellular process in which the protein mentioned under [2)] is involved. [4)] CCPI, clathrin-coated pit internalization; TGN, trans-Golgi network. [5)] Literature references are indicated by numbers in between parentheses (*1*: Gotthardt *et al.*, 2000; *2*: He *et al.*, 2002; *3*: Mishra *et al.*, 2002; *4*: Matter *et al.*, 1992; *5*: Matter *et al.*, 1994; *6*: Hoppe and Joiner, 2002; *7*: Li *et al.*, 2000; *8*: Kyttala *et al.*, 2004; *9*: Rapoport *et al.*, 1998; *10*: Trommsdorff *et al.*, 1998; *11*: Petersen *et al.*, 2003; *12*: Rader *et al.*, 2000; *13*: Oleinikov *et al.*, 2000; *14*: Hjälm *et al.*, 1996; *15*: Larsson *et al.*, 2003; *16*: Stockinger *et al.*, 2000). If no reference is indicated, then the motif is deduced from the amino acid sequence file with the accession number mentioned in Table 1. "?" indicates that the function is deduced by homology of the motif to that in family members for which the function has been described. [6)] Ø represents an amino acid with a bulky hydrophobic side chain. [7)] Indicates the sequence from H^{37} to V^{81}. P^{46} participates in the sequence to which MegBP and Ankra bind (H^{37} to V^{81}). The motif PPPSP in which P^{204} participates may not be an SH3 binding site, since it appears to function in a phosphorylation-dependent interaction with axin in LRP5/6/Arrow (Tamai *et al.*, 2004). The latter receptors are not included in this study since they lack a recognizable NPXY motif. [8)] Most likely this motif does not function in degradation since it is located distantly from the M(X9)G motif. [9)] Currently, many different PDZ-domain containing proteins are found to bind to the ICD of LRP2. PSD95 is indicated here as a representative of this class of proteins. [10)] The sequence of the non-spliced variant of LRP8 (Stockinger *et al.*, 2000) is used here. [11)] Indicates the sequence from A^{159} to K^{217}. [12)] This motif fulfills requirements for interaction with the sorting machinery, since acidic residues are located two and four residues N-terminal of L^{155}, however, it may also function as a class I PDZ-domain binding motif (Sheng and Sala, 2001).

cytoplasmic domain. It is currently not known if and how the mosaic structure affects the function of the individual motifs. However, the x-ray crystallization study by Stolt *et al.* (Stolt *et al.*, 2003; for review see Jeon and Blacklow, 2005) of the adaptor protein Dab1 and a peptide of human LRP8, containing the sequence TKSMNFDNPVYRKT that includes the sequence of the motifs II, III, IVa, and V (Figure 1), suggests that interjacent sequences may affect the function of the individual motif; receptor groups that contain interjacent sequences in these four motifs harbor a second clathrin-coated pit internalization motif, motif IVb (*i.e.* LRP, LRP2, LRP8, and Ce LRP1). Therefore, disruption of a linear chain of the motifs II, III, IVa, and V may modulate the function of each of the individual sequence motifs. In this respect the LpRs resemble the LDLR, VLDLR, and VGRs.

With respect to the complete ICD, the insect receptors are similarly organized as the LDLR, VLDLR, and VGR groups, apart from both their extreme N- and C-terminal sequence. As indicated above, in the N-terminal sequence of the domain they resemble the LRP2 group. Interestingly, at their C-terminus, they resemble the LRP2 group of receptors

and Ce LRP1 by the presence of a PDZ-domain interacting sequence (LRP2: SEVc, LpR: DYVc; Figure 1; Table 2). The LRP2 receptors would be recognized by a class I PDZ domain and the insect receptor group by a class III domain. PDZ domains are modular protein interaction domains that bind in a sequence specific fashion to short C-terminal or internal peptides. Proteins may contain multiple copies of this domain that has a scaffold function in the cell. These proteins are involved in localized signaling or trafficking (Sheng and Sala, 2001). The PDZ-binding motif in LRP2 appears not to be involved in the apical cell sorting (Takeda *et al.*, 2003). Most likely, the binding specificity of the PDZ domain is not confined solely to the last three residues, and may involve additional residues, in particular those that are present in the extensions of the two receptor groups. LRP2 has been shown to interact with PDZ-domain containing proteins like PSD95 (Gotthardt *et al.*, 2000; Larsson *et al.*, 2003). The role of the interaction of LRP2 with PDZ-domain containing protein is still unknown. For the insect receptors, no studies have yet been performed to identify proteins that specifically interact with their PDZ-binding motif.

Conclusions

In this study, we show that LDLR family members can be classified based on the amino acid sequence of their non-recurring C-terminal domains. This classification resembles earlier classifications that were done on the basis of nucleotide or amino acid sequences, in combination with the number of cysteine-rich ligand-binding repeats in the N-terminal ligand-binding domain (see Sappington and Raikhel, 1998a,b; Willnow, 1999; Herz and Bock, 2002), and shows that the non-repeat containing domains harbor information that uniquely defines properties that belong to a particular receptor or group of receptors. Based on the data presented in this study, the insect lipoprotein receptors, LpRs, characterized as potentially binding lipophorin and/or expressed in fat body tissue, constitute a separate group of LDLR family members. LpRs contain an OLGD, and considering the overall organization of the three C-terminal domains, the insect receptors resemble most that of the LDLR, VLDLR, and VGR groups. Interestingly, however, in the sequence positions where the LpR group deviates from the other three groups, the insect receptors resemble the LRP2 group, as indicated by the absence of a P residue in the TMD, a Y-residue in the first position of motif I, the presence of an extended C-terminus in the ICD, and the presence of a C-terminal PDZ-domain interacting motif.

These group-specific sequences may constitute the basis for the identification of other potential lipophorin receptors (*e.g.* in genome sequence projects). Accordingly, the identified LDLR family member in the *Drosophila* genome project with accession number RE38584 may not be a lipophorin receptor, or may not be expressed in the fat body. Consequently, the ones with the accession numbers GH26833 and EAA03681 (from *A.*

gambiae) (included in this study) may bind lipophorin or may be expressed in fat body. Based on the same reasons, the earlier identified VLDLR homologue from *A. aegypti* (Cheon *et al.*, 2001) also appears to be a typical lipophorin receptor.

The unique combination of sequence motifs in the three domains of the insect receptors suggests that they perform a unique function or set of functions. The(se) function(s) may not be completely exclusive, as the motifs are also recovered in other family members. On the other hand, we cannot exclude that unique functions appear in one specific receptor, by the specific mosaic clustering of motifs that occur in different groups of receptors. Nevertheless, based on the data presented here, the insect receptors can be called typical LDLR family members. The insect family members used in this study appear less homologous among themselves, than the family members within the other groups. This may well be related to the evolutionary distance between the insect species from which the receptors derive, and does not necessarily reflect differences in function. Knowledge about the function of the unique interjacent and flanking sequences in the ICD acquired for each individual LpR will be helpful to describe the insect receptor function in general. Hence, the unique ligand recycling property of Lm LpR in mammalian (Van Hoof *et al.*, 2002) and possibly in insect cells (Van Hoof *et al.*, 2003, 2005a), may be defined by the function of extended C-terminal sequence and PDZ-domain interaction motif, possibly associated with that of other (interjacent) motifs.

Acknowledgements

Part of the work in the authors' laboratory was supported by the Earth and Life Sciences Foundation (ALW) that is subsidized by the Netherlands Organization of Scientific Research (NWO; grant number ALW-809.38.003 to DVH), and the Department of Biology of Utrecht University, The Netherlands.

Chapter 6

Summary and General Discussion

Most major lipoproteins across the animal kingdom are constituted by members of the LLTP family. The homologous LLTP members apoLp-II/I and apoB constitute lipoproteins in insects and vertebrates, respectively. Insect apoLp-II/I lipoproteins differ from vertebrate apoB lipoproteins in the major neutral lipid (DAG rather than TAG), the amount of lipids bound, as well as the proteolytic cleavage of apoLp-II/I, unlike apoB. These differences combine in the recognition of two outstanding events during the biogenesis of apoLp-II/I into a lipoprotein: 1) cleavage of apoLp-II/I into apoLp-I and apoLp-II (Weers *et al.*, 1992), and 2) lipidation to a high-density lipoprotein (Weers *et al.*, 1993), in contrast to the formation of very low-density lipoproteins from apoB in mammals. From this comparative perspective, this thesis focuses on the biosynthesis of insect lipoprotein, and extends to the analysis of the diversity and evolution of LLTP as well as their receptors from the LDLR family.

Despite the noted differences among insect apoLp-II/I and vertebrate apoB lipoproteins, the results in this thesis highlight the structural similarity of apoB and apoLp-II/I, also relative to other LLTP, by virtue of (1) the sequence homology among their LLT modules by which both are recognized as members of the apoB-like LLTP subfamily (Chapter 4), (2) the sequence homology outside of the LLT module (Chapter 4), (3) the amphipathic clusters predicted (Chapters 3 and 4), (4) the capacity of apoLp-II/I LLT module to constitute a lipid-binding cavity structure (Chapter 2), like apoB (Segrest *et al.*, 1999; Richardson *et al.*, 2005) and other LLTP (Anderson *et al.*, 1998; Mann *et al.*, 1999; Thompson and Banaszak, 2002), (5) the location of the apoLp-II/I cleavage site in a region of the LLT module that has not been preserved among LLTP (Chapters 2 and 4), and (6) the ability of apoLp-II/I to constitute a lipophorin, independent of its cleavage, like apoB (Chapters 2 and 3).

The cleavage of apoLp-II/I was shown to be performed by an insect furin, *i.e.* a furin-like proprotein convertase, given the repression of apoLp-II/I cleavage by 1) mutagenesis of the consensus substrate sequence for furin, and 2) competitive inhibition of apoLp-II/I cleavage by a furin substrate inhibitor (Chapter 2). The cleavage of apoLp-II/I is a marked, yet not essential event in initial insect lipoprotein biogenesis, as indicated by the ability of locust fat body (Chapter 2) and a recombinant apoLp-II/I expression system (Chapter 3) to constitute high-density lipophorin from uncleaved apoLp-II/I. However, the conservation of apoLp-II/I cleavage in all insects characterized to date emphasizes the importance of this post-translational modification. Several suggestions on the adaptive value of apoLp-II/I cleavage have been discussed in Chapter 2, *e.g.*

enhancing the flexibility of the LLT module lipid-binding cavity to allow for further lipidation of HDLp to LDLp, modifying/enabling endocytic uptake by lipophorin receptor, and enabling distinct functions for apoLp-I and apoLp-II in hemolymph coagulation. Thus, the cleavage of apoLp-II/I into apoLp-I and apoLp-II is not essential to lipoprotein formation, but is proposed to represent a molecular adaptation for which the functional value remains to be established.

The lipidation of apoLp-II/I-derived proteins to an HDLp was studied in Chapter 3. Insect apoLp-II/I were predicted to contain clusters enriched in either amphipathic α-helices or amphipathic β-strands, organized along the protein as N-α_1-β-α_2-C, reminiscent of a truncated form of apoB that has a clustering corresponding to N-α_1-β_1-α_2-β_2-α_3-C. Expression of C-terminal truncation variants of *L. migratoria* apoLp-II/I in a recombinant lipoprotein expression system demonstrates that the region corresponding to the β cluster enables apoLp-II/I to bind sufficient lipids to constitute a high-density lipoprotein, while the α_1 cluster and N-terminal part of the β cluster only do not suffice for buoyant lipoprotein formation. Similarly, the β_1 cluster of apoB has previously been demonstrated to enable sufficient lipid binding for buoyant lipoprotein biogenesis (Yao *et al.*, 1991; Spring *et al.*, 1992; Dashti *et al.*, 2002; Shelness *et al.*, 2003; Manchekar *et al.*, 2004). Whereas insect lipoprotein appears to bind lipids via the predicted amphipathic β cluster in apoLp-II/I, our results also suggest that insect MTP facilitates the acquisition of lipids by apoLp-II/I, as co-expression of *D. melanogaster* MTP enhanced apoLp-II/I secretion as well as lipidation (Chapter 3). The importance of MTP in insect lipoprotein biosynthesis *in vivo* remains to be established. Strikingly, MTP is essential for apoB lipoprotein biosynthesis in mammals and likely all vertebrates (Marza *et al.*, 2005; for reviews see Mahmood Hussain *et al.*, 2001; Shelness and Sellers, 2001; Olofsson and Borén, 2005; Shoulders and Shelness, 2005; Shelness and Ledford, 2005) and also enhances the secretion of another LLTP, Vtg, in invertebrates as well as vertebrates (Shibata *et al.*, 2003; Sellers *et al.*, 2005). Thus, insect lipoprotein formation from apoLp-II/I seems to depend on the same major elements as lipoprotein biogenesis from apoB in mammals, *i.e.* amphipathic β cluster sequence as well as MTP.

In view of the homology of apoLp-II/I to apoB, lipidation likely preceeds cleavage of apoLp-II/I into apoLp-I and apoLp-II. The lipidation of apoB starts co-translationally in the rough endoplasmic reticulum and is completed post-translationally in the smooth endoplasmic reticulum and/or cis-Golgi network, possibly by fusion with an intralumenal lipid droplet (for reviews see Mahmood Hussain *et al.*, 2001; Shelness and Sellers, 2001; Olofsson and Borén, 2005). Upon its translocation, the N-terminal LLT module of apoB folds and associates with MTP. This interaction enables lipid acquisition of the nascent apoB polypeptide and accordingly prevents misfolding and aggregation of C-terminal amphipathic structures, and the resulting intracellular retention and degradation of apoB. Based on the presence of an LLT module and similar amphipathic clusters in apoLp-II/I

(Babin *et al.*, 1999; Mann *et al.*, 1999; Chapters 2, 3 and 4) as well as the likely involvement of MTP in insect lipoprotein assembly (Chapter 3), the lipidation of apoLp-II/I probably starts co-translationally as well. However, cleavage by furin homologs is performed late in the secretory pathway, mainly in the trans-Golgi network (for review see Molloy *et al.*, 1999). Therefore, lipidation likely starts prior to apoLp-II/I cleavage. Cleavage of apoLp-II/I prior to any lipidation might result in the aggregation of apoLp-II/I proteins or the parting of apoLp-I and apoLp-II, and hence impairment of lipoprotein biosynthesis. Nonetheless, lipidation to the mature HDLp may be completed only following cleavage of apoLp-II/I into apoLp-I and apoLp-II, possibly even following secretion of lipophorin in some species (Prasad *et al.*, 1986; Capurro and De Bianchi, 1990), so the order of events in lipoprotein biogenesis remains to be established experimentally.

The present findings on *L. migratoria* HDLp biosynthesis likely apply to the biogenesis of lipophorin in other insects as well. ApoLp-II/I characterized in other insects from distinct orders also contain a consensus substrate sequence for furin as the probable site for precursor cleavage (Kutty *et al.*, 1996; Sundermeyer *et al.*, 1996), and have an LLT module as well as an amphipathic clustering corresponding to N-α_1-β-α_2-C (Chapters 3 and 4). Moreover, MTP has been identified in several insect species (Sellers *et al.*, 2003; Chapter 4). Unlike investigations in other insect species (Capurro and De Bianchi, 1990; Weers *et al.*, 1992; Van Heusden *et al.*, 1998; Chapter 2), for *Manduca sexta* the larval fat body was reported to secrete lipophorin as a very high-density lipophorin (VHDLp), that was further lipidated into the mature HDLp at the intestine (Prasad *et al.*, 1986). This apparently distinct lipidation process may reflect the lipid-poor status of the fat bodies, due to developmental stage or starvation of the animals used. Interestingly, lipophorin biosynthesis may not only be mediated by the fat body (Prasad *et al.*, 1986; Capurro and De Bianchi, 1990; Weers *et al.*, 1992; Van Heusden *et al.*, 1998), as apoLp-II/I is also expressed in the locust eye (Bogerd *et al.*, 2000). It will be of interest to investigate possible diversity in the degree and localization of lipidation, dependent on the species, developmental stage and tissue of lipophorin biogenesis.

The involvement of similar elements in the biogenesis of insect apoLp-II/I lipophorin as well as vertebrate apoB lipoproteins, *i.e.* MTP and amphipathic regions, implies an ancient metazoan origin for current mechanisms of lipoprotein biogenesis. To gain understanding of the origin of lipoproteins and the mechanisms for their biogenesis, the evolutionary origin and diversification of the family of LLTP were studied (Chapter 4). Three subfamilies of LLTP were recognized. ApoLp-II/I and apoB were found to belong to the same subfamily, named apoB-like LLTP, identified by phylogenetic reconstruction of conserved segments in the LLT module and in addition supported by the presence of a motif in these sequences only, *i.e.* pfam06448 (Marchler-Bauer *et al.*, 2005), as well as the prediction of a pronounced clustering of amphipathic β-strands. The other two recognized

subfamilies are Vtg-like LLTP and the MTP group. LLTP could only be identified confidently in metazoan phyla, that were found to have distinct subsets of LLTP. For example, the nematode *C. elegans* appears to lack apoB-like LLTP, and whereas mammals lack Vtg-like LLTP, some teleost fish encode over ten different Vtg (Buisine *et al.*, 2002). Strikingly, two novel LLTP were identified in insects, *i.e.* the previously characterized melanin-enhancing protein (MEP; Lee *et al.*, 2000) that is a Vtg-like LLTP, and the apoLp-II/I-related protein (ARP) that is an apoB-like LLTP. The function of ARP remains to be established, yet its extended amphipathic β cluster and similarity to apoLp-II/I suggest a function in lipid transport.

The early evolution and origin of LLTP subfamilies and their domains, *i.e.* the LLT module, vWF-D module and amphipathic clusters, could not be established due to the high sequence divergence as well as the present lack of LLTP sequences from early-diverged metazoans. Current efforts on additional metazoan genomes, particularly from the ancient metazoan phyla of Cnidaria and Porifera, may provide further views on the early evolution of LLTP lipoproteins and their mechanism of biogenesis.

Strikingly, the major yolk protein in decapodan Crustacea is not a Vtg-like but an apoB-like LLTP, exemplifying the functional flexibility of LLTP. A wealth of LLTP functionality may still await discovery, as is illustrated by the very recent discoveries that apoLp-II/I transports morphogen proteins during early embryonal development of *D. melanogaster* (Panakova *et al.*, 2005), and that mammalian MTP facilitates lipid antigen presentation (Brozovic *et al.*, 2004; Dougan *et al.*, 2005). Although the early evolution and divergence of LLTP subfamilies and domains could not be traced, this is the first study to demonstrate convincingly the close phylogenetic relation between apoB and apoLp-II/I, and to address the functional divergence of LLTP from an evolutionary perspective.

Whereas LLTP family members constitute the structural protein basis for lipoproteins and their biogenesis, metazoans rely on members of the LDLR family, a homologous set of lipoprotein receptors, for the cellular uptake of specific LLTP family members. The present classification of newly discovered LDLR family members is based on the amount of cysteine-rich repeats in the ligand-binding domain (see Willnow, 1999; Herz and Bock, 2002; Jeon and Blacklow, 2005). However, the number of repeats is known to expand with relatively high frequency (Sappington and Raikhel, 1998b), and accordingly may not be an appropriate tool to properly group novel invertebrate LDLR family members. In Chapter 5, we show that classification of LDLR family members based on the amino acid sequence of their non-repetitive C-terminal domains results in the grouping of vertebrate LDLR family members similar to previous classifications. In addition, the novel method of classification characterizes the insect lipoprotein receptors (LpR), *i.e.* insect family members that potentially bind lipophorin and/or are expressed in fat body tissue, as a separate group. Therefore, the non-recurring domains of LDLR family members harbor

information that uniquely defines properties that belong to a particular receptor or group of receptors. The unique combination of sequence features in LpR was reviewed in relation to the functional differences disclosed between insect and mammalian lipoprotein receptors.

Thus, the major results of this thesis are: (1) the non-exchangeable apolipoprotein of insect lipoproteins, apoLp-II/I, is cleaved by an insect furin, (2) cleavage is not essential to high-density lipoprotein formation, as uncleaved apoLp-II/I can be secreted and constitute a high-density lipoprotein, (3) apoLp-II/I requires a region predicted to be enriched in amphipathic β-strands to constitute a high-density lipophorin, (4) MTP stimulates lipophorin secretion and lipidation in a recombinant expression system, (5) apoLp-II/I is an apoB-like LLTP, based on sequence homology and secondary structure prediction, and (6) LDLR family members can be classified using their non-recurring domains, which modulate LDLR function via specific sequence motifs.

Accordingly, the work presented in this thesis demonstrates that lipoprotein biogenesis and lipid transport in mammals and insects are governed by common molecular elements and mechanisms. Modification of these elements, by the interplay of evolution and biogenesis, results in compositional and functional differences among lipoproteins. This perspective yields several intriguing potential explanations to specific properties of lipoproteins and lipid transport in insects, as illustrated by the examples below:

- A single modification in the intracellular lipid metabolism pathways may explain why DAG is the most common neutral lipid in insect lipoproteins, rather than TAG as in mammalian apoB lipoproteins. In mammals, most of the TAG used for apoB lipoprotein assembly originates from cytosolic TAG stores. However, TAG does not readily enter into membranes, unlike DAG. Accordingly, lipoprotein assembly in the secretory pathway depends on the lipolysis of cytosolic TAG into DAG and fatty acid. DAG permeates the ER membrane and is re-esterified to TAG on the lumenal side of the membrane by a diacylglycerol acyltransferase (DGAT) (Owen *et al.*, 1997; Abo-Hashema *et al.*, 1999; for reviews see Gibbons *et al.*, 2000, 2004). In this way, TAG is believed to become available for the lipidation of apoB. In insects, the conversion of HDLp into LDLp and likely the biosynthesis of HDLp also depend on the lipolysis of cytosolic TAG stores (Van der Horst and Ryan, 2004). Accordingly, the presence of DAG in insect lipoprotein may reflect the absence of sufficient DGAT activity at the luminal side of membranes, be it endoplasmic reticulum or plasma membrane. Thus, the lack of luminal DGAT activity offers a mechanistic explanation for the mobilization of DAG, rather than TAG, onto insect lipoproteins.

- Unlike mammalian apoB lipoproteins, lipophorin functions as a reusable lipid transporter. The hemolymph factor LTP enables lipophorin reusability, as it performs its extracellular lipid reloading by facilitating the redistribution of lipids among lipoproteins or the plasma membrane of cells (for reviews see Ryan and Van der Horst, 2000; Van der Horst and Ryan, 2004). Based on the analogous activity to LTP and the secretion of dMTP by recombinant insect cells (Chapter 3), MTP is speculated to be a component of the LTP complex. The secretion of MTP, rather than its intracellular retention, therefore could enable the reusability of insect lipoprotein.

- ApoLp-III facilitates the binding of additional lipids by lipophorin upon enhanced lipid mobilization as occurs during sustained insect flight (for reviews see Weers and Ryan, 2003; Van der Horst and Ryan, 2004). ApoLp-III is constituted by amphipathic α-helices. In apoB, these secondary structures have been proposed to stabilize the constituted lipoproteins upon changes in lipid content, by reversibly associating or dissociating from the lipid core (Segrest *et al.*, 2001; Wang *et al.*, 2003; Wang and Small, 2004). ApoLp-II/I, and accordingly apoLp-I and –II, were predicted to contain a relative small amount of amphipathic α-helices, as compared to apoB (Chapters 3 and 4). Therefore it is proposed that apoLp-III is a molecular adaptation that compensates for the relatively small amount of amphipathic α-helices present in apoLp-I and –II, as compared to apoB, hence stabilizing lipophorin upon changes in lipid content.

These examples illustrate that the recognition of the ancient metazoan origin of the structural elements and mechanisms of lipoprotein biosynthesis, evidenced in this thesis, sets the stage for progressing from recognizing differences between mammalian and insect lipid transport, towards understanding their molecular origin and functional relevance. Moreover, the present findings provide a conceptual framework for further comparative studies on lipoproteins of invertebrate animals, and an impetus to the integration of knowledge from different organizational levels, *e.g.* molecules, cells, tissues, organisms, species, and phyla, to arrive at a completer understanding of mechanisms of lipid transport. In addition to their intrinsic value, comparative studies on lipoproteins and lipid transport may provide novel perspectives on human lipoprotein function and on the treatment of lipid disorders.

References

Abascal, F., Zardoya, R., and Posada, D. (2005) ProtTest: Selection of best-fit models of protein evolution. *Bioinformatics* **21,** 2104-2105.

Abdu, U., Davis, C., Khalaila, I., and Sagi, A. (2002) The vitellogenin cDNA of *Cherax quadricarinatus* encodes a lipoprotein with calcium binding ability, and its expression is induced following the removal of the androgenic gland in a sexually plastic system. *Gen. Comp. Endocrinol.* **127,** 263-272.

Abo-Hashema, K.A., Cake, M.H., Power, G.W., and Clarke, D. (1999) Evidence for triacylglycerol synthesis in the lumen of microsomes via a lipolysis-esterification pathway involving carnitine acyltransferases. J. Biol. Chem. **274,** 35577-35582.

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

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25,** 3389-3402.

Amdam, G.V., Aase, A.L., Seehuus, S.C., Kim Fondrk, M., Norberg, K., and Hartfelder, K. (2005) Social reversal of immunosenescence in honey bee workers. *Exp. Gerontol.* **40,** 939-947.

Andersen, O.M., and Petersen, H.H. (2003) New light on a long-known protein family. *ChemBioChem* **4,** 1137-1146.

Anderson, T.A., Levitt, D.G., and Banaszak, L.J. (1998) The structural basis of lipid interactions in lipovitellin, a soluble lipoprotein. *Structure* **6,** 895-909.

Assmann, G., and Nofer, J.R. (2003) Atheroprotective effects of high-density lipoproteins. *Annu. Rev. Med.* **54,** 321-341.

Avarre, J.C., Michelis, R., Tietz, A., Lubzens, E. (2003) Relationship between vitellogenin and vitellin in a marine shrimp (Penaeus semisulcatus) and molecular characterization of vitellogenin complementary DNAs. *Biol. Reprod.* **69**, 355-364.

Babin, P.J., Bogerd, J., Kooiman, F.P., Van Marrewijk, W.J.A., and Van der Horst, D.J. (1999) Apolipophorin II/I, apolipoprotein B, vitellogenin, and microsomal triglyceride transfer protein genes are derived from a common ancestor. *J. Mol. Evol.* **49,** 150-160.

Barwig, B. (1985) Isolation and characterization of plasma coagulogen (PC) of the cockroach *Leucophaea maderae* (Blattaria). *J. Comp. Physiol. [B]*. **155,** 135–143.

Beenakkers, A.M.Th., Van der Horst, D.J., and Van Marrewijk, W.J.A. (1985) Insect lipids and lipoproteins, and their role in physiological processes. *Prog. Lipid Res.* **24,** 19-67.

Blumenthal, T., Squire, M., Kirtland, S., Cane, J., Donegan, M., Spieth, J., and Sharrock, W. (1984) Cloning of a yolk protein gene family from Caenorhabditis elegans. *J. Mol. Biol.* **174**, 1-18.

Bogerd, J., Babin, P.J., Kooiman, F.P., André, M., Ballagny, C., Van Marrewijk, W.J.A., and Van der Horst, D.J. (2000) Molecular characterization and gene expression in the eye of the apolipophorin II/I precursor from *Locusta migratoria*. *J. Comp. Neurol.* **427,** 546-558.

Bownes, M., Lineruth, K., and Mauchline, D. (1991) Egg production and fertility in *Drosophila* depend upon the number of yolk-protein gene copies. *Mol. Gen. Genet.* **228,** 324-327.

Bownes, M. (1992) Why is there sequence similarity between insect yolk proteins and vertebrate lipases? *J. Lipid Res.* **33,** 777-790.

Brandt, B.W., Zwaan, B.J., Beekman, M., Westendorp, R.G., and Slagboom, P.E. (2005) Shuttling between species for pathways of lifespan regulation: A central role for the vitellogenin gene family? *Bioessays* **27,** 339-346.

Brehélin, M. (1979) Hemolymph coagulation in *Locusta migratoria*: evidence for a functional equivalent of fibrinogen. *Comp. Biochem. Physiol. B.* **62,** 329–334.

Brozovic, S., Nagaishi, T., Yoshida, M., Betz, S., Salas, A., Chen, D., Kaser, A., Glickman, J., Kuo, T., Little, A., Morrison, J., Corazza, N., Kim, J.Y., Colgan, S.P., Young, S.G., Exley, M., and Blumberg, R.S. (2004) CD1d function is regulated by microsomal triglyceride transfer protein. *Nat. Med.* **10,** 535-539.

Buisine, N., Trichet, V., and Wolff, J. (2002) Complex evolution of vitellogenin genes in salmonid fishes. *Mol. Genet. Genomics* **268,** 535-542.

C. elegans Sequencing Consortium (1998) Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* **282,** 2012-2018.

Capurro, M. de L., and De Bianchi, A.G. (1990) Larval *Musca domestica* lipophorin biosynthesis. *Comp. Biochem. Physiol.* **97B,** 655-659.

Chapman, M.J. (1980) Animal lipoproteins: chemistry, structure, and comparative aspects. *J. Lipid Res.* **21,** 789-853.

Chen, J.S., and Raikhel, A.S. (1996) Subunit cleavage of mosquito pro-vitellogenin by a subtilisin-like convertase. *Proc. Natl. Acad. Sci. USA.* **93,** 6186–6190.

Chen, J.S., Sappington, T.W., and Raikhel, A.S. (1997) Extensive sequence conservation among insect, nematode, and vertebrate vitellogenins reveals ancient common ancestry. *J. Mol. Evol.* **44,** 440-451.

Chen, S.H., Yang, C.Y., Chen, P.F., Setzer, D., Tanimura, M., Li, W.H., Gotto, A.M. Jr. and Chan, L. (1986) The complete cDNA and amino acid sequence of human apolipoprotein B-100. *J. Biol. Chem.* **261,** 12918-12921.

Cheon, H.M., Seo, S.J., Sun, J., Sappington, T.W., and Raikhel, A.S. (2001) Molecular characterization of the VLDL receptor homolog mediating binding of lipophorin in oocyte of the mosquito Aedes aegypti. *Insect Biochem. Mol. Biol.* **31,** 753-760.

Chino, H., Downer, R.G.H., Wyatt, G.R., and Gilbert, L.I. (1981) Lipophorins, a major class of lipoproteins of insect hemolymph. *Insect Biochem.* **11,** 491.

Chino, H., Hirayama, Y., Kiyomoto, Y., Downer, R.G.H., and Takahashi, K. (1987) Spontaneous aggregation of locust lipophorin during hemolymph collection. *Insect Biochem.* **17,** 89–97.

Cieplik, M., Klenk, H.D., and Garten, W. (1998) Identification and characterization of *Spodoptera frugiperda* furin: a thermostable subtilisin-like endopeptidase. *Biol. Chem.* **379,** 1433–1440.

Clatworthy, A.E., Stockinger, W., Christie, R.H., Schneider, W.J., Nimpf, J., Hyman, B.T., and Rebeck, G.W. (1999) Expression and alternate splicing of apolipoprotein E receptor 2 in brain. *Neuroscience* **90,** 903-911.

Dantuma, N.P., Van Marrewijk, W.J.A., Wynne, H.J., and Van der Horst, D.J. (1996) Interaction of an insect lipoprotein with its binding site at the fat body. *J. Lipid Res.* **37,** 1345-1355.

Dantuma, N.P., Potters, M., De Winther, M.P., Tensen, C.P., Kooiman, F.P., Bogerd, J., and Van der Horst, D.J. (1999) An insect homolog of the vertebrate very low density lipoprotein receptor mediates endocytosis of lipophorins. *J. Lipid Res.* **40,** 973-978.

Dashti, N., Gandhi, M., Liu, X., Lin, X., and Segrest, J.P. (2002) The N-terminal 1000 residues of apolipoprotein B associate with microsomal triglyceride transfer protein to create a lipid transfer pocket required for lipoprotein assembly. *Biochemistry* **41,** 6978-6987.

Davidson, N.O., and Shelness, G.S. (2000) Apolipoprotein B: mRNA editing, lipoprotein assembly, and presecretory degradation. *Annu. Rev. Nutr.* **20,** 169-193.

Davidson, W.S., and Silva, R.A. (2005) Apolipoprotein structural organization in high density lipoproteins: belts, bundles, hinges and hairpins. *Curr. Opin. Lipidol.* 16(3): 295-300.

De Bie, I., Savaria, D., Roebroek, A.J.M., Day, R., Lazure, C., Van de Ven, W.J.M., and Seidah, N.G. (1995) Processing specificity and biosynthesis of the *Drosophila melanogaster* convertases dfurin1, dfurin1-CRR, dfurin1-X, and dfurin2. *J. Biol. Chem.* **270,** 1020–1028.

Djordjevic, J.T., Waterkeyn, J.G., Hennessy, K.L., and Stanley, K.K. (2000) Role of phosphorylation of the cytoplasmic domain of the alpha(2)-macroglobulin receptor (LRP). *Cell Biol. Int.* **24,** 599-610.

Donnell, D.M. (2004) Vitellogenin of the parasitoid wasp, *Encarsia formosa* (Hymenoptera: Aphelinidae): gene organization and differential use by members of the genus. *Insect Biochem. Mol. Biol.* **34,** 951-961.

Dougan, S.K., Salas, A., Rava, P., Agyemang, A., Kaser, A., Morrison, J., Khurana, A., Kronenberg, M., Johnson, C., Exley, M., Hussain, M.M., and Blumberg, R.S. (2005) Microsomal triglyceride transfer protein lipidation and control of CD1d on antigen-presenting cells. *J. Exp. Med.* **202,** 529-539.

Duvic, B., and Brehélin, M. (1998) Two major proteins from locust plasma are involved in coagulation and are specifically precipitated by laminarin, a β-1,3-glucan. *Insect Biochem. Mol. Biol.* **28,** 959– 967.

Epton, M.J., Dilworth, R.J., Smith, W., Hart, B.J., and Thomas, W.R. (1999) High-molecular-weight allergens of the house dust mite: an apolipophorin-like cDNA has sequence identity with the major M-177 allergen and the IgE-binding peptide fragments Mag1 and Mag3. *Int. Arch. Allergy Immunol.* **120,** 185-191.

Frayn, K.N. (2003) Lipoprotein metabolism. In: *Metabolic regulation. A human perspective.* 2nd ed. Blackwell Science, Oxford, U.K. 253-280.

Gellissen, G. (1983) Lipophorin as the plasma coagulogen in *Locusta migratoria. Naturwissenschaften.* **70,** 45–46.

Gerber, D., Sal-Man, N., and Shai, Y. (2004) Two motifs within a transmembrane domain, one for homodimerization and the other for heterodimerization. *J. Biol. Chem.* **279,** 21177-21182.

Gibbons, G.F., Islam, K., and Pease, R.J. (2000) Mobilisation of triacylglycerol stores. *Biochim. Biophys. Acta.* **1483,** 37-57.

Gibbons, G.F., Wiggins, D., Brown, A.M., and Hebbachi, A.M. (2004) Synthesis and function of hepatic very-low-density lipoprotein. *Biochem. Soc. Trans.* **32,** 59-64.

Goldstein, J.L., Brown, M.S., Anderson, R.G., Russell, D.W., Schneider, W.J. (1985) Receptor-mediated endocytosis: concepts emerging from the LDL receptor system. *Ann. Rev. Cell Biol.* **1,** 1-39.

Gotthardt, M., Trommsdorff, M., Nevitt, M.F., Shelton, J., Richardson, J.A., Stockinger, W., Nimpf, J., and Herz, J. (2000) Interactions of the low density lipoprotein receptor gene family with cytosolic adaptor and scaffold proteins suggest diverse biological functions in cellular communication and signal transduction. *J. Biol. Chem.* **275,** 25616-25624.

Graham, D.L., Knott, T.J., Jones, T.C., Pease, R.J., Pullinger, C.R., and Scott, J. (1991) Carboxyl-terminal truncation of apolipoprotein B results in gradual loss of the ability to form buoyant lipoproteins in cultured human and rat liver cell lines. *Biochemistry* **30,** 5616-5621.

Gretch, D.G., Sturley, S.L., Wang, L., Lipton, B.A., Dunning, A., Grunwald, K.A., Wetterau, J.R., Yao, Z., Talmud, P., and Attie, A.D. (1996) The amino terminus of apolipoprotein B is necessary but not sufficient for microsomal triglyceride transfer protein responsiveness. *J. Biol. Chem.* **271,** 8682-8691.

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

Guindon, S., Lethiec, F., Duroux, P., and Gascuel, O. (2005) PHYML Online--a web server for fast maximum likelihood-based phylogenetic inference. *Nucleic Acids Res.* **33,** W557-559.

Hall, M., Wang, R., Van Antwerpen, R., Sottrup-Jensen, L., and Söderhäll, K.. (1999) The crayfish plasma clotting protein: a vitellogenin-related protein responsible for clot formation in crustacean blood. *Proc. Natl. Acad. Sci. USA* **96,** 1965-1970.

He, G., Gupta, S., Yi, M., Michaely, P., Hobbs, H.H., and Cohen, J.C. (2002) ARH is a modular adaptor protein that interacts with the LDL receptor, clathrin, and AP-2. *J. Biol. Chem.* **277,** 44044-44049.

Henikoff, S., and Henikoff, J.G. (1993) Performance evaluation of amino acid substitution matrices. *Proteins* **17,** 49-61.

Hennig, L. (1999) WinGene/WinPep: user-friendly software for the analysis of amino acid sequences. *Biotechniques* **26,** 1170-1172.

Henrich, S., Cameron, A., Bourenkov, G.P., Kiefersauer, R., Huber, R., Lindberg, I., Bode, W., and Than, M.E. (2003) The crystal structure of the proprotein processing proteinase furin explains its stringent specificity. *Nat. Struct. Biol.* **10,** 520–526.

Heras, H., and Pollero, R.J. (2002) Lipoproteins from plasma and perivelline fluid of the apple snail *Pomacea canaliculata*. *Biocell* **26,** 111-118.

Herz, J., and Strickland, D.K. (2001) LRP: a multifunctional scavenger and signaling receptor. *J. Clin. Invest.* **108,** 779-784.

Herz, J., and Bock, H.H. (2002) Lipoprotein receptors in the nervous system. *Ann. Rev. Biochem.* **71,** 405-434.

Hessa, T., Monne, M., and von Heijne, G. (2003) Stop-transfer efficiency of marginally hydrophobic segments depends on the length of the carboxy-terminal tail. *EMBO Rep.* **4,** 178-183.

Hevonoja, T., Pentikainen, M.O., Hyvonen, M.T., Kovanen, P.T., and Ala-Korpela, M. (2000) Structure of low density lipoprotein (LDL) particles: basis for understanding molecular changes in modified LDL. *Biochim. Biophys. Acta* **1488,** 189-210.

Hjälm, G., Murray, E., Crumley, G., Harazim, W., Lundgren, S., Onyango, I., Ek, B., Larsson, M., Juhlin, C., Hellman, P., Davis, H., Akerstrom, G., Rask, L., and Morse, B. (1996) Cloning and sequencing of human gp330, a Ca^{2+}-binding receptor with potential intracellular signaling properties. *Eur. J. Biochem.* **239,** 132-137.

Hofmann, K., and Stoffel, W. (1993) TMbase - A database of membrane spanning proteins segments. *Biol. Chem. Hoppe-Seyler* **374,** 166.

Holt, R.A., Subramanian, G.M., Halpern, A., Sutton, G.G., Charlab, R., Nusskern, D.R., Wincker, P., Clark, A.G., Ribeiro, J.M., Wides, R., Salzberg, S.L., Loftus, B., Yandell, M., Majoros, W.H., Rusch, D.B., Lai, Z., Kraft, C.L., Abril, J.F., Anthouard, V., Arensburger, P., Atkinson, P.W., Baden, H., de Berardinis, V., Baldwin, D., Benes, V., Biedler, J., Blass, C., Bolanos, R., Boscus, D., Barnstead, M., Cai, S., Center, A., Chatuverdi, K., Christophides, G.K., Chrystal, M.A., Clamp, M., Cravchik, A., Curwen, V., Dana, A., Delcher, A., Dew, I., Evans, C. A., Flanigan, M., Grundschober-Freimoser, A., Friedli, L., Gu, Z., Guan, P., Guigo, R., Hillenmeyer, M.E., Hladun, S. L., Hogan, J.R., Hong, Y.S., Hoover, J., Jaillon, O., Ke, Z., Kodira, C., Kokoza, E., Koutsos, A., Letunic, I., Levitsky, A., Liang, Y., Lin, J.J., Lobo, N.F., Lopez, J.R., Malek, J.A., McIntosh, T.C., Meister, S., Miller, J., Mobarry, C., Mongin, E., Murphy, S. D., O'Brochta, D.A., Pfannkoch, C., Qi, R., Regier, M.A., Remington, K., Shao, H., Sharakhova, M.V., Sitter, C.D., Shetty, J., Smith, T.J., Strong, R., Sun, J., Thomasova, D., Ton, L.Q., Topalis, P., Tu, Z., Unger, M.F., Walenz, B., Wang, A., Wang, J., Wang, M., Wang, X., Woodford, K.J., Wortman, J.R., Wu, M., Yao, A., Zdobnov, E.M., Zhang, H., Zhao, Q., Zhao, S., Zhu, S.C., Zhimulev, I., Coluzzi, M., della Torre, A., Roth, C.W., Louis, C., Kalush, F., Mural, R.J., Myers, E.W., Adams, M. D., Smith, H.O., Broder, S., Gardner, M.J., Fraser, C.M., Birney, E., Bork, P., Brey, P. T., Venter, J.C., Weissenbach, J., Kafatos, F.C., Collins, F.H., and Hoffman, S.L. (2002) The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science* **298,** 129-149.

Hoppe, H.C., and Joiner, K.A. (2000) Cytoplasmic tail motifs mediate endoplasmic reticulum localization and export of transmembrane reporters in the protozoan parasite Toxoplasma gondii. *Cell Microbiol.* **2,** 569-578.

Huelsenbeck, J.P., and Ronquist, F. (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754-755.

Iijima, H., Miyazawa, M., Sakai, J., Magoori, K., Ito, M.R., Suzuki, H., Nose, M., Kawarabayasi, Y., and Yamamoto, T.T. (1998) Expression and characterization of a very low density lipoprotein receptor variant lacking the O-linked sugar region generated by alternative splicing. *J. Biochem. (Tokyo)* **124,** 747-755.

Jean, F., Stella, K., Thomas, L., Liu, G., Xiang, Y., Reason, A.J., and Thomas, G. (1998) α1-Antitrypsin Portland, a bioengineered serpin highly selective for furin: application as an antipathogenic agent. *Proc. Natl. Acad. Sci. USA.* **95,** 7293–7298.

Jeon, H., and Blacklow, S.C. (2005) Structure and physiologic function of the low-density lipoprotein receptor. *Annu. Rev. Biochem.* **74,** 535-562.

Journet, A.M., Saffaripour, S., and Wagner, D.D. (1993) Requirement for both D domains of the propolypeptide in von Willebrand factor multimerization and storage. *Thromb. Haemost.* **70,** 1053-1057.

Kato, Y., Tokishita, S., Ohta, T., and Yamagata, H. (2004) A vitellogenin chain containing a superoxide dismutase-like domain is the major component of yolk proteins in cladoceran crustacean *Daphnia magna*. *Gene* **334,** 157-165.

Kawooya, J.K., Osir, E.O., and Law, J.H. (1988) Uptake of the major hemolymph lipoprotein and its transformation in the insect egg. *J. Biol. Chem.* **263,** 8740-8747.

Kawooya, J.K., and Law, J.H. (1988) Role of lipophorin in lipid transport to the insect egg. *J. Biol. Chem.* **263,** 8748-8753.

Kawooya, J.K., Wells, M.A., and Law, J.H. (1989) A strategy for solubilizing delipidated apolipoprotein with lysophosphatidylcholine and reconstitution with phosphatidylcholine. *Biochemistry.* **28,** 6658–6667.

Killian, J.A., and von Heijne, G. (2000) How proteins adapt to a membrane-water interface. *Trends Biochem. Sci.* **25,** 429-434.

Killian, J.A. (2003) Synthetic peptides as models for intrinsic membrane proteins. *FEBS L.* **555,** 134-138.

Kingsley, D.M., Kozarsky, K.F., Segal, M., and Krieger, M. (1986) Three types of low density lipoprotein receptor-deficient mutant have pleiotropic defects in the synthesis of N-linked, O-linked, and lipid-linked carbohydrate chains. *J. Cell Biol.* **102,** 1576-1585.

Kollman, J.M., and Quispe, J. (2005) The 17Å structure of the 420 kDa lobster clottable protein by single particle reconstruction from cryoelectron micrographs. *J. Struct. Biol.* **151,** 306-314.

Korschineck, I., Ziegler, S., Breuss, J., Lang, I., Lorenz, M., Kaun, C., Ambros, P.F., and Binder, B.R. (2001) Identification of a novel exon in apolipoprotein E receptor 2 leading to alternatively spliced mRNAs found in cells of the vascular wall but not in neuronal tissue. *J. Biol. Chem.* **276,** 13192-13197.

Krieger, M., and Herz, J. (1994) Structures and functions of multiligand lipoprotein receptors: macrophage scavenger receptors and LDL receptor-related protein (LRP). *Annu. Rev. Biochem.* **63,** 601-637.

Kung, S.Y., Chan, S.M., Hui, J.H., Tsang, W.S., Mak, A., and He, J.G. (2004) Vitellogenesis in the sand shrimp, *Metapenaeus ensis*: the contribution from the hepatopancreas-specific vitellogenin gene (MeVg2). *Biol Reprod.* **71,** 863-70.

Kutty, R.K., Kutty, G., Kambadur, R., Duncan, T., Koonin, E.V., Rodriguez, I.R., Odenwald, W.F., and Wiggert, B. (1996) Molecular characterization and developmental expression of a retinoid- and fatty acid-binding glycoprotein from *Drosophila*. A putative lipophorin. *J. Biol. Chem.* **271,** 20641-20649.

Kyttala, A., Ihrke, G., Vesa, J., Schell, M.J., and Luzio, J.P. (2004) Two motifs target Batten disease protein CLN3 to lysosomes in transfected nonneuronal and neuronal cells. *Mol. Biol. Cell* **15,** 1313-1323.

Lamberg, A., Jauhiainen, M., Metso, J., Ehnholm, C., Shoulders, C., Scott, J., Pihlajaniemi T, and Kivirikko, K.I. (1996) The role of protein disulphide isomerase in the microsomal triacylglycerol transfer protein does not reside in its isomerase activity. *Biochem. J.* **315,** 533-536.

Larsson, M., Hjalm, G., Sakwe, A.M., Engstrom, A., Hoglund, A.S., Larsson, E., Robinson, R.C., Sundberg, C., and Rask, L. (2003) Selective interaction of megalin with postsynaptic density-95 (PSD-95)-like membrane-associated guanylate kinase (MAGUK) proteins. *Biochem. J.* **373,** 381-391.

Lee, C.S., Han, J.H., Kim, B.S., Lee, S.M., Hwang, J.S., Kang, S.W., Lee, B.H., and Kim, H.R. (2003) Wax moth, Galleria mellonella, high density lipophorin receptor: alternative splicing, tissue-specific expression, and developmental regulation. *Insect Biochem. Mol. Biol.* **33,** 761-771.

Lee, K.M., Lee, K.Y., Choi, H.W., Cho, M.Y., Kwon, T.H., Kawabata, S., and Lee, B.L. (2000) Activated phenoloxidase from *Tenebrio molitor* larvae enhances the synthesis of melanin by using a vitellogenin-like protein in the presence of dopamine. *Eur. J. Biochem.* **267,** 3695-3703.

Leiper, J.M., Bayliss, J.D., Pease, R.J., Brett, D.J., Scott, J., and Shoulders, C.C. (1994) Microsomal triglyceride transfer protein, the abetalipoproteinemia gene product, mediates the secretion of apolipoprotein B-containing lipoproteins from heterologous cells. *J. Biol. Chem.* **269,** 21951-21954.

Li, C.M., Presley, J.B., Zhang, X., Dashti, N., Chung, B.H., Medeiros, N.E., Guidry, C., and Curcio, C.A. (2005a) Retina expresses microsomal triglyceride transfer protein: implications for age-related maculopathy. *J. Lipid Res.* 46, 628-640.

Li, C.M., Chung, B.H., Presley, J.B., Malek, G., Zhang, X., Dashti, N., Li, L., Chen, J., Bradley, K., Kruth, H.S., and Curcio, C.A. (2005b) Lipoprotein-like particles and cholesteryl esters in human Bruch's membrane: initial characterization. *Invest. Ophthalmol. Vis. Sci.* **46,** 2576-2586.

Li, D., Scherfer, C., Korayem, A.M., Zhao, Z., Schmidt, O., and Theopold, U. (2002) Insect hemolymph clotting: evidence for interaction between the coagulation system and the prophenoloxidase activating cascade. *Insect Biochem. Mol. Biol.* **32,** 919–928.

Li, Y., Marzolo, M.P., van Kerkhof, P., Strous, G.J., and Bu, G. (2000) The YXXL motif, but not the two NPXY motifs, serves as the dominant endocytosis signal for low density lipoprotein receptor-related protein. *J. Biol. Chem.* **275,** 17187-17194.

Lim, E.H., Teo, B.Y., Lam, T.J., and Ding, J.L. (2001) Sequence analysis of a fish vitellogenin cDNA with a large phosvitin domain. *Gene* **277,** 175-186.

Lynch, M., and Conery, J.S. (2000) The evolutionary fate and consequences of duplicate genes. *Science* **290,** 1151-1155.

Ma, G., Hay, D., Li, D., Asgari, S., and Schmidt, O. (2006) Recognition and inactivation of LPS by lipophorin particles. *Dev. Comp. Immunol.* **in press**

Maddison, D.R., and Schulz, K.-S. (eds.) (2004) The Tree of Life Web Project. Internet address: http://tolweb.org.

Madsen, E.M., Lindegaard, M.L., Andersen, C.B., Damm, P., and Nielsen, L.B. (2004) Human placenta secretes apolipoprotein B-100-containing lipoproteins. *J. Biol. Chem.* **279,** 55271-55276.

Magrane, J., Casaroli-Marano, R.P., Reina, M., Gafvels, M., and Vilaro, S. (1999) The role of O-linked sugars in determining the very low density lipoprotein receptor stability or release from the cell. *FEBS Lett.* **451,** 56-62.

Mahmood Hussain, M., Kedees, M.H., Singh, K., Athar, H., and Jamali, N.Z. (2001) Signposts in the assembly of chylomicrons. *Front. Biosci.* **6,** D320–D331.

Mahmood Hussain, M., Shi, J., and Dreizen, P. (2003) Microsomal triglyceride transfer protein and its role in apoB-lipoprotein assembly. *J. Lipid Res.* **44,** 22-32.

Mak, A.S., Choi, C.L., Tiu, S.H., Hui, J.H., He, J.G., Tobe, S.S., and Chan, S.M. (2005) Vitellogenesis in the red crab *Charybdis feriatus*: Hepatopancreas-specific expression and farnesoic acid stimulation of vitellogenin gene expression. *Mol. Reprod. Dev.* **70,** 288-300.

Manchekar, M., Richardson, P.E., Forte, T.M., Datta, G., Segrest, J.P., and Dashti, N. (2004) Apolipoprotein B-containing lipoprotein particle assembly: Lipid capacity of the nascent lipoprotein particle. *J. Biol. Chem*. **279,** 39757-39766.

Mann, C.J., Anderson, T.A., Read, J., Chester, S..A., Harrison, G.B., Kochl, S., Ritchie, P.J., Bradbury, P., Hussain, F.S., Amey, J., Vanloo, B., Rosseneu, M., Infante, R., Hancock, J.M., Levitt, D.G., Banaszak, L.J., Scott, J., and Shoulders, C.C. (1999) The structure of vitellogenin provides a molecular model for the assembly and secretion of atherogenic lipoproteins. *J. Mol. Biol.* **285,** 391-408.

Marchler-Bauer, A., Anderson, J.B., Cherukuri, P.F., DeWeese-Scott, C., Geer, L.Y., Gwadz, M., He, S., Hurwitz, D.I., Jackson, J.D., Ke, Z., Lanczycki, C., Liebert, C.A., Liu, C., Lu, F., Marchler, G.H., Mullokandov, M., Shoemaker, B.A., Simonyan, V., Song, J.S., Thiessen, P.A., Yamashita, R.A., Yin, J.J., Zhang, D., and Bryant, S.H. CDD: a Conserved Domain Database for protein classification. *Nucleic Acids Res*. **33,** D192-196.

Marsh, J.B. (1968) Isolation and composition of a low density lipoprotein from the eggs of *Arbacia punctulata*. *Biol. Bull.* **135,** 193-199.

Martensen, P.M., Oka, K., Christensen, L., Rettenberger, P.M., Petersen, H.H., Christensen, A., Chan, L., Heegaard, C.W., and Andreasen, P.A. (1997) Breast carcinoma epithelial cells express a very low-density lipoprotein receptor variant lacking the O-linked glycosylation domain encoded by exon 16, but with full binding activity for serine proteinase/serpin complexes and Mr-40,000 receptor-associated protein. *Eur. J. Biochem.* **248,** 583-591.

Martins, I.J., and Redgrave, T.G. (2004) Obesity and post-prandial lipid metabolism. Feast or famine? *J. Nutr. Biochem.* **15,** 130-141.

Marza, E., Barthe, C., André, M., Villeneuve, L., Hélou, C. and Babin, P.J. (2005) Developmental expression and nutritional regulation of a zebrafish gene homologous to mammalian microsomal triglyceride transfer protein large subunit. *Dev. Dyn.* **232,** 506-518.

Matsumoto, T., Nakamura, A.M., Mori, K., and Kayano, T. (2003) Molecular characterization of a cDNA encoding putative vitellogenin from the Pacific oyster *Crassostrea gigas*. *Zoolog. Sci.* **20,** 37-42.

Matter, K., Hunziker, W., Mellman, I. (1992) Basolateral sorting of LDL receptor in MDCK cells: the cytoplasmic domain contains two tyrosine-dependent targeting determinants. *Cell* **71,** 741-753.

Matter, K., Yamamoto, E.M., and Mellman, I. (1994) Structural requirements and sequence motifs for polarized sorting and endocytosis of LDL and Fc receptors in MDCK cells. *J. Cell Biol.* **126,** 991-1004.

Maxfield, F.R., and McGraw, T.E. (2004) Endocytic recycling. *Nature Rev. Mol. Cell Biol.* **5,** 121-132.

May, P., Reddy, Y.K., and Herz, J. (2002) Proteolytic processing of low density lipoprotein receptor-related protein mediates regulated release of its intracellular domain. *J. Biol. Chem.* **277,** 18736-18743.

May, P., and Herz, J. (2003) LDL receptor-related proteins in neurodevelopment. *Traffic* **4,** 291-301.

May, P., Bock, H.H., and Herz, J. (2003) Integration of endocytosis and signal transduction by lipoprotein receptors. *Science's STKE* **2003,** PE12.

May, P., Bock, H.H., Nimpf, J., and Herz, J. (2003) Differential glycosylation regulates processing of lipoprotein receptors by gamma-secretase. *J. Biol. Chem.* **278,** 37386-37392.

May, P., Herz, J., and Bock, H.H. (2005) Molecular mechanisms of lipoprotein receptor signalling. *Cell. Mol. Life Sci.* **62,** 2325-2338.

McLeod, R.S., Wang, Y., Wang, S., Rusinol, A., Links, P., and Yao, Z. (1996) Apolipoprotein B sequence requirements for hepatic very low density lipoprotein assembly. Evidence that hydrophobic sequences within apolipoprotein B48 mediate lipid recruitment. *J. Biol. Chem.* **271,** 18445-18455.

Melman, L., Geuze, H.J., Li, Y., McCormick, L.M., Van Kerkhof, P., Strous, G.J., Schwartz, A.L., and Bu, G. (2002) Proteasome regulates the delivery of LDL receptor-related protein into the degradation pathway. *Mol. Biol. Cell* **13,** 3325-3335.

Mishra, S.K., Keyel, P.A., Hawryluk, M.J., Agostinelli, N.R., Watkins, S.C., and Traub, L.M. (2002) Disabled-2 exhibits the properties of a cargo-selective endocytic clathrin adaptor. *EMBO J.* **21,** 4915-4926.

Molloy, S.S., Anderson, E.D., Jean, F., and Thomas, G. (1999) Bicycling the furin pathway: from TGN localization to pathogen activation and embryogenesis. *Trends Cell Biol.* **9,** 28–35.

Monné, M., and von Heijne, G. (2001) Effects of 'hydrophobic mismatch' on the location of transmembrane helices in the ER membrane. *FEBS Letters* **496,** 96-100.

Montgomery, R., Conway, T.W., Spector, A.A., and Chapel, D. (1996) Lipoproteins In: *Biochemistry. A case-oriented approach*. 4th ed. Mosby-Year Book Inc., St. Louis. 356-383.

Montorzi, M., Falchuk, K.H., and Vallee, B.L. (1995) Vitellogenin and lipovitellin: zinc proteins of *Xenopus laevis* oocytes. *Biochemistry* **34,** 10851-10858.

Nagai, M., Meerloo, T., Takeda, T., and Farquhar, M.G. (2003) The adaptor protein ARH escorts megalin to and through endosomes. *Mol. Biol. Cell* **14,** 4984-4996.

Nakamura, Y., Yamamoto, M., and Kumamaru, E. (1998) A variant very low density lipoprotein receptor lacking 84 base pairs of O-linked sugar domain in the human brain myelin. *Brain Res.* **793,** 47-53.

Nardelli, D., van het Schip, F.D., Gerber-Huber, S., Haefliger, J.A., Gruber, M., Ab, G., and Wahli, W. (1987) Comparison of the organization and fine structure of a chicken and a *Xenopus laevis* vitellogenin gene. *J. Biol. Chem.* **262,** 15377-15385.

Notredame, C., Higgins, D., and Heringa, J. (2000) T-Coffee: a novel algorithm for multiple sequence alignment. *J. Mol. Biol.* **302,** 205–217.

Okuno, A., Yang, W.J., Jayasankar, V., Saido-Sakanaka, H., Huong, D.T.T., Jasmani, S., Atmomarsono, M., Subramoniam, T., Tsutsui, N., Ohira, T., Kawazoe, I., Aida, K., and Wilder, M.N. (2002) Deduced primary structure of vitellogenin in the giant freshwater prawn, *Macrobrachium rosenbergii*, and yolk processing during ovarian maturation. *J. Exp. Zool.* **292,** 417-29.

Oleinikov, A.V., Zhao, J., and Makker, S.P. (2000) Cytosolic adaptor protein Dab2 is an intracellular ligand of endocytic receptor gp600/megalin. *Biochem. J.* **347,** 613-621.

Olofsson, S.O., and Borén, J. (2005) Apolipoprotein B: a clinically important apolipoprotein which assembles atherogenic lipoproteins and promotes the development of atherosclerosis. *J. Intern. Med.* **258,** 395-410.

Owen, M.R., Corstorphine, C.C., and Zammit, V.A. (1997) Overt and latent activities of diacylglycerol acytransferase in rat liver microsomes: possible roles in very-low-density lipoprotein triacylglycerol secretion. *Biochem. J.* **323,** 17-21.

Page, R.D.M. (1996) TREEVIEW: An application to display phylogenetic trees on personal computers. *Comp. Appl. Biosci.* **12,** 357-358.

Panakova, D., Sprong, H., Marois, E., Thiele, C., and Eaton, S. (2005) Lipoprotein particles are required for Hedgehog and Wingless signalling. *Nature* **435,** 58-65.

Patrie, K.M., Drescher, A.J., Goyal, M., Wiggins, R.C., and Margolis, B. (2001) The membrane-associated guanylate kinase protein MAGI-1 binds megalin and is present in glomerular podocytes. *J. Am. Soc. Nephrol.* **12,** 667-677.

Perez-Vilar, J., and Hill, R.L. (1999) The structure and assembly of secreted mucins. J. Biol. Chem. **274,** 31751-31754.

Petersen, H.H., Hilpert, J., Militz, D., Zandler, V., Jacobsen, C., Roebroek, A.J., and Willnow, T.E. (2003) Functional interaction of megalin with the megalinbinding protein (MegBP), a novel tetratrico peptide repeat-containing adaptor molecule. *J. Cell Sci.* **116,** 453-461.

Poirot, O., O'Toole, E., and Notredame, C. (2003) Tcoffee@igs: A web server for computing, evaluating and combining multiple sequence alignments. *Nucleic Acids Res.* **31,** 3503-3506.

Prasad, S.V., Fernando-Warnakulasuriya, G.J.P., Sumida, M., Law, J.H., and Wells, M.A. (1986) Lipoprotein biosynthesis in the larvae of the tobacco hornworm, *Manduca sexta*. *J. Biol. Chem.* **36,** 17174-17176.

Pruitt, K.D., Tatusova, T., and Maglott, D.R. (2005) NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res*. **33,** D501-D504.

Raag, R., Appelt, K., Xuong, N.H., and Banaszak, L. (1988) Structure of the lamprey yolk lipid-protein complex lipovitellin-phosvitin at 2.8 Å resolution. *J. Mol. Biol.* **200,** 553-569.

Rader, K., Orlando, R.A., Lou, X., and Farquhar, M.G. (2000) Characterization of ANKRA, a novel ankyrin repeat protein that interacts with the cytoplasmic domain of megalin. *J. Am. Soc. Nephrol.* **11,** 2167-2178.

Rapoport, I., Miyazaki, M., Boll, W., Duckworth, B., Cantley, L.C., Shoelson, S., and Kirchhausen, T. (1997) Regulatory interactions in the recognition of endocytic sorting signals by AP-2 complexes. *EMBO J.* **16,** 2240-2250.

Rapoport, I., Chen, Y.C., Cupers, P., Shoelson, S.E., and Kirchhausen, T. (1998) Dileucine-based sorting signals bind to the beta chain of AP-1 at a site distinct and regulated differently from the tyrosine-based motif-binding site. *EMBO J.* **17,** 2148-2155.

Raviv, S., Parnes, S., Segall, C., Davis, C., and Sagi, A. (2006) Complete sequence of *Litopenaeus vannamei* (Crustacea: Decapoda) vitellogenin cDNA and its expression in endocrinologically induced sub-adult females. *Gen. Comp. Endocrinol.* **145,** 39-50.

Read, J., Anderson, T.A., Ritchie, P.J., Vanloo, B., Amey, J., Levitt, D., Rosseneu, M., Scott, J., and Shoulders, C.C. (2000) A mechanism of membrane neutral lipid acquisition by the microsomal triglyceride transfer protein. *J. Biol. Chem.* **275,** 30372-30377.

Rehemtulla, A., and Kaufman, R.J. (1992) Preferred sequence requirements for cleavage of pro-von Willebrand factor by propeptide-processing enzymes. *Blood.* **79,** 2349–2355.

Richardson, P.E., Manchekar, M., Dashti, N., Jones, M.K., Beigneux, A., Young, S.G., Harvey, S.C., and Segrest, J.P. (2005) Assembly of lipoprotein particles containing apolipoprotein-B: structural model for the nascent lipoprotein particle. *Biophys. J.* **88,** 2789-2800.

Robbs, S.L., Ryan, R.O., Schmidt, J.O., Keim, P.S., and Law, J.H. (1985) Lipophorin of the larval honeybee, *Apis mellifera* L. *J. Lipid Res.* **26,** 241-247.

Rockwell, N.C., Krysan, D.J., Komiyama, T., and Fuller, R.S. (2002) Precursor processing by kex2/furin proteases. *Chem. Rev.* **102,** 4525–4548.

Rodenburg, K.W., and Van der Horst, D.J. (2005) Lipoprotein-mediated lipid transport in insects: Analogy to the mammalian lipid carrier system and novel concepts for the functioning of LDL receptor family members. *Biochim. Biophys. Acta* **1736,** 10-29.

Roebroek, A.J.M., Pauli, I.G.L., Zhang, Y., and Van de Ven, W.J.M. (1991) cDNA sequence of a *Drosophila melanogaster* gene, Dfur1, encoding a protein structurally related to the subtilisin-like proprotein processing enzyme furin. *FEBS Lett.* **289,** 133–137.

Roebroek, A.J.M., Creemers, J.W.M., Pauli, I.G.L., Kurzik-Dumke, U., Rentrop, M., Gateff, E.A.F., Leunissen, J.A.M., and Van de Ven, W.J.M. (1992) Cloning and functional expression of Dfurin2, a subtilisin-like proprotein processing enzyme of *Drosophila melanogaster* with multiple repeats of a cysteine motif. *J. Biol. Chem.* **267,** 17208–17215.

Roebroek, A.J.M., Creemers, J.W.M., Ayoubi, T.A., and Van de Ven, W.J.M. (1994) Furin-mediated proprotein processing activity: involvement of negatively charged amino acid residues in the substrate binding region. *Biochimie.* **76,** 210–216.

Romano, M., Rosanova, P., Anteo, C., and Limatola, E. (2004) Vertebrate yolk proteins: a review. *Mol. Reprod. Dev.* **69,** 109-116.

Ryan, R.O., Van der Horst, D.J. (2000) Lipid transport biochemistry and its role in energy. (2000) *Annu Rev Entomol.* **45**, 233-60.

Šali, A., and Blundell, T.L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234,** 779–815.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual.* 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sappington, T.W., Kokoza, V.A., Cho, W.L., and Raikhel, A.S. (1996) Molecular characterization of the mosquito vitellogenin receptor reveals unexpected high homology to the *Drosophila* yolk protein receptor. *Proc. Natl. Acad. Sci. USA* **93,** 8934-8939.

Sappington, T.W., and Raikhel, A.S. (1998a) Molecular characteristics of insect vitellogenins and vitellogenin receptors. *Insect Biochem. Mol. Biol.* **28,** 277-300.

Sappington, T.W., and Raikhel, A.S. (1998b) Ligand-binding domains in vitellogenin receptors and other LDL-receptor family members share a common ancestral ordering of cysteine-rich repeats. *J. Mol. Evol.* **46,** 476-487.

Sappington, T.W. (2002) The major yolk proteins of higher Diptera are homologs of a class of minor yolk proteins in lepidoptera. *J. Mol. Evol.* **55,** 470-475.

Schechter, I., and Berger, A. (1967) On the size of the active site in proteases. *Biochem. Biophys. Res. Commun.* **27,** 157–162.

Schenk, S., Harris, J.R., Hoeger, U. (2006) A discoidal lipoprotein from the coelomic fluid of the polychaete *Nereis virens*. *Comp. Biochem. Physiol. B. Biochem Mol Biol.* **143,** 236-243.

Schneider, W.J. (1996) Vitellogenin receptors: oocyte-specific members of the low-density lipoprotein receptor supergene family. *Int. Rev. Cytol.* **166,** 103-137.

Schneider, W.J., and Nimpf, J. (2003) LDL receptor relatives at the crossroad of endocytosis and signaling. *Cell. Mol. Life Science* **60,** 892-903.

Schulz, T.K.F., Van der Horst, D.J., Amesz, H., Voorma, H.O., and Beenakkers, A.M.Th. (1987). Monoclonal antibodies specific for apoproteins of lipophorins from the migratory locust. *Arch. Insect Pysiol. Biochem.* **6,** 97-107.

Segrest, J.P., Jones, M.K., Mishra, V.K., Anantharamaiah, G.M., and Garber, D.W. (1994) ApoB-100 has a pentapartite structure composed of three amphipathic α-helical domains alternating with two amphipathic β-strand domains. Detection by the computer program LOCATE. *Arterioscler. Thromb.* **14,** 1674-1685.

Segrest, J.P., Jones, M.K., Mishra, V.K., Pierotti, V., Young, S.H., Borén, J., Innerarity, T.L., and Dashti, N. (1998) Apolipoprotein B-100: conservation of lipid-associating amphipathic secondary structural motifs in nine species of vertebrates. *J. Lipid Res.* **39,** 85-102.

Segrest, J.P., Jones, M.K., and Dashti, N. (1999) N-terminal domain of apolipoprotein B has structural homology to lipovitellin and microsomal triglyceride transfer protein: a "lipid pocket" model for self-assembly of apoB-containing lipoprotein particles. *J. Lipid Res.* **40,** 1401-1416.

Segrest, J.P., Jones, M.K., De Loof, H., and Dashti, N. (2001) Structure of apolipoprotein B-100 in low density lipoproteins. *J. Lipid Res.* **42,** 1346-1367.

Sellers, J.A., Hou, L., Athar, H., Hussain, M.M., and Shelness, G.S. (2003) A *Drosophila* microsomal triglyceride transfer protein homolog promotes the assembly and secretion of human apolipoprotein B. Implications for human and insect transport and metabolism. *J. Biol. Chem.* **278,** 20367-73.

Sellers, J.A., Hou, L., Schoenberg, D.R., Batistuzzo de Medeiros, S.R., Wahli, W., and Shelness, G.S. (2005) Microsomal triglyceride transfer protein promotes the secretion of *Xenopus laevis* Vitellogenin A1. *J. Biol. Chem.* **280,** 13902-13905.

Seo, S.J., Cheon, H.M., Sun, J., Sappington, T.W., and Raikhel, A.S. (2003) Tissue- and stage-specific expression of two lipophorin receptor variants with seven and eight ligand-binding repeats in the adult mosquito. *J. Biol. Chem.* **278,** 41954-41962.

Shapiro, J.P., Keim, P.S., and Law, J.H. (1984) Structural studies on lipophorin, an insect lipoprotein. *J. Biol. Chem.* **259,** 3680-3685.

Sharp, D., Blinderman, L., Combs, K.A., Kienzle, B., Ricci, B., Wager-Smith, K., Gil, C.M., Turck, C.W., Bouma, M.-E., Rader, D.J., Aggerbeck, L.P., Gregg, R.E., Gordon, D.A., and Wetterau, J.R. (1993) Cloning and gene defects in microsomal triglyceride transfer protein associated with abetalipoproteinaemia. *Nature* **365,** 65-69.

Sharrock, W.J., Rosenwasser, T.A., Gould, J., Knott, J., Hussey, D., Gordon, J.I., and Banaszak, L. (1992) Sequence of lamprey vitellogenin. Implications for the lipovitellin crystal structure. *J. Mol. Biol.* **226,** 903-907.

Shelness, G.S., and Sellers, J.A. (2001) Very-low-density lipoprotein assembly and secretion. *Curr. Opin. Lipidol.* **12,** 151-157.

Shelness, G.S., Hou, L., Ledford, A.S., Parks, J.S., and Weinberg, R.B. (2003) Identification of the lipoprotein initiating domain of apolipoprotein B. *J. Biol. Chem.* **278,** 44702–44707.

Shelness, G.S., and Ledford, A.S. (2005) Evolution and mechanism of apolipoprotein B-containing lipoprotein assembly. *Curr. Opin. Lipidol.* **16,** 325-332.

Sheng, M., and Sala, C. (2001) PDZ domains and the organization of supramolecular complexes. *Annu. Rev. Neurosci.* **24,** 1-29.

Shibata, Y., Branicky, R., Landaverde, I. O., and Hekimi, S. (2003) Redox regulation of germline and vulval development in *Caenorhabditis elegans*. *Science* **302,** 1779-1782.

Shoulders, C.C., Brett, D.J., Bayliss, J.D., Narcisi, T.M.E., Jarmuz, A., Grantham, T.T., Leoni, P.R.D., Bhattacharya, S., Pease, R.J., Cullen, P.M., Levi, S., Byfield, P.G.H., Purkiss, P., and Scott, J. (1993) Abetalipoproteinemia is caused by defects of the gene encoding the 97 kDa subunit of a microsomal triglyceride transfer protein. *Hum. Mol. Genet.* **2,** 2109-2116.

Shoulders, C.C., and Shelness, G.S. (2005) Current biology of MTP: implications for selective inhibition. *Curr. Top. Med. Chem.* **5,** 283-300.

Sigrist, C.J.A., Cerutti L., Hulo N., Gattiker A., Falquet L., Pagni M., Bairoch A., and Bucher, P. (2002) PROSITE: a documented database using patterns and profiles as motif descriptors. *Brief. Bioinform.* **3,** 265-274.

Smolenaars, M.M.W., Kasperaitis, M.A.M., Richardson, P.E., Rodenburg, K.W., and Van der Horst, D.J. (2005) *Supplemental data to* J. Lipid Res., vol. 46, pp. 412–421. Internet address: http://www.jlr.org/cgi/content/full/M400374-JLR200/DC2.

Soulages, J.L., and Wells, M.A. (1994) Lipophorin: the structure of an insect lipoprotein and its role in lipid transport in insects. *Adv. Protein Chem.* **45,** 371-415.

Spieth, J., Nettleton, M., Zucker-Aprison, E., Lea, K., and Blumenthal, T. (1991) Vitellogenin motifs conserved in nematodes and vertebrates. *J. Mol. Evol.* **32,** 429-438.

Spring, D.J., Chen-Liu, L.W., Chatterton, J.E., Elovson, J., Schumaker, V.N. (1992) Lipoprotein assembly. Apolipoprotein B size determines lipoprotein core circumference. *J. Biol. Chem.* **267,** 14839-14845.

Stapleton, M., Liao, G., Brokstein, P., Hong, L., Carninci, P., Shiraki, T., Hayashizaki, Y., Champe, M., Pacleb, J., Wan, K., Yu, C., Carlson, J., George, R., Celniker, S., and Rubin, G.M. (2002) The Drosophila gene collection: identification of putative full-length cDNAs for 70% of *D. melanogaster* genes. *Genome Res.* **12,** 1294-1300.

Stockinger, W., Brandes, C., Fasching, D., Hermann, M., Gotthardt, M., Herz, J., Schneider, W.J., and Nimpf, J. (2000) The reelin receptor ApoER2 recruits JNK-interacting proteins-1 and -2. *J. Biol. Chem.* **275,** 25625-25632.

Stockinger, W., Sailler, B., Strasser, V., Recheis, B., Fasching, D., Kahr, L., Schneider, W.J., and Nimpf, J. (2002) The PX-domain protein SNX17 interacts with members of the LDL receptor family and modulates endocytosis of the LDL receptor. *EMBO J.* **21,** 4259-4267.

Stolt, P.C., Jeon, H., Song, H.K., Herz, J., Eck, M.J., and Blacklow, S.C. (2003) Origins of peptide selectivity and phosphoinositide binding revealed by structures of disabled-1 PTB domain complexes. *Structure (Camb)* **11,** 569-579.

Strickland, D.K., and Ranganathan, S. (2003) Diverse role of LDL receptor-related protein in the clearance of proteases and in signaling. *J. Thromb. Haemost.* **1,** 1663-1670.

Strimmer, K., and von Haeseler, A. (1996) Quartet puzzling: A quartet maximum likelihood method for reconstructing tree topologies. *Mol. Biol. Evol.* **13,** 964-969.

Sun, J., Hiraoka, T., Dittmer, N.T., Cho, K.H., and Raikhel, A.S. (2000) Lipophorin as a yolk protein precursor in the mosquito, *Aedes aegypti*. *Insect Bioch. Mol. Biol.* **30,** 1161-1171.

Sun, X.M., and Soutar, A.K. (1999) Expression *in vitro* of alternatively spliced variants of the messenger RNA for human apolipoprotein E receptor-2 identified in human tissues by ribonuclease protection assays. *Eur. J. Biochem.* **262,** 230-239.

Sundermeyer, K., Hendricks, J.K., Prasad, S.V., and Wells, M.A. (1996) The precursor protein of the structural apolipoproteins of lipophorin: cDNA and deduced amino acid sequence. *Insect Biochem. Mol. Biol.* **26,** 735–738.

Takeda, T., Yamazaki, H., and Farquhar, M.G. (2003) Identification of an apical sorting determinant in the cytoplasmic tail of megalin. *Am. J. Physiol. Cell Physiol.* **284,** C1105-1113.

Tamai, K., Zeng, X., Liu, C., Zhang, X., Harada, Y., Chang, Z., and He, X. (2004) A mechanism for Wnt coreceptor activation. *Mol. Cell* **13,** 149-156.

Taskinen, M.R. (2005) Type 2 diabetes as a lipid disorder. *Curr. Mol. Med.* **5,** 297-308.

Taylor, N.A., Van de Ven, W.J.M., and Creemers, J.W.M. (2003) Curbing activation: proprotein convertases in homeostasis and pathology. *FASEB J.* **17,** 1215–1227.

Terasawa, Y., Cases, S.J., Wong, J.S., Jamil, H., Jothi, S., Traber, M.G., Packer, L., Gordon, D.A., Hamilton, R.L., and Farese, R.V. Jr. (1999) Apolipoprotein B-related gene expression and ultrastructural characteristics of lipoprotein secretion in mouse yolk sac during embryonic development. *J. Lipid Res.* **40,**1967-1977.

Theopold, U., and Schmidt, O. (1997) *Helix pomatia* lectin and annexin V, two molecular probes for insect microparticles: possible involvement in hemolymph coagulation. *J. Insect Physiol.* **43,** 667–674. 46.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22,** 4673–4680.

Thompson, J.R., and Banaszak, L.J. (2002) Lipid-protein interactions in lipovitellin. *Biochemistry* **41,** 9398-9409.

Tidball, J.G. (1981) Lipoprotein secretion during the formation of the axial skeleton in *Leptogorgia virgulata* (Cnidaria: Gorgonacea). *Cell Tissue Res.* **219,** 327-338.

Tong, Y., Shan, T., Poh, Y.K., Yan, T., Wang, H., Lam, S.H., and Gong, Z. (2004) Molecular cloning of zebrafish and medaka vitellogenin genes and comparison of their expression in response to 17β-estradiol. *Gene* **328,** 25-36.

Trewitt, P.M., Heilmann, L.J., Degrugillier, S.S., and Kumaran, A.K. (1992) The boll weevil vitellogenin gene: nucleotide sequence, structure, and evolutionary relationship to nematode and vertebrate vitellogenin genes. *J. Mol. Evol.* **34,** 478-492.

Trommsdorff, M., Gotthardt, M., Hiesberger, T., Shelton, J., Stockinger, W., Nimpf, J., Hammer, R.E., Richardson, J.A., and Herz, J. (1999) Reeler/Disabled-like disruption of neuronal migration in knockout mice lacking the VLDL receptor and ApoE receptor 2. *Cell* **97,** 689-701.

Tsang, W.S., Quackenbush, L.S., Chow, B.K., Tiu, S.H., He, J.G., and Chan, S.M. (2003) Organization of the shrimp vitellogenin gene: evidence of multiple genes and tissue specific expression by the ovary and hepatopancreas. *Gene* **303,** 99-109.

Tsutsui, N., Kawazoe, I., Ohira, T., Jasmani, S., Yang, W.J., Wilder, M.N., and Aida, K. (2000) Molecular characterization of a cDNA encoding vitellogenin and its expression in the hepatopancreas and ovary during vitellogenesis in the kuruma prawn, *Penaeus japonicus*. *Zool. Sci.* **17,** 651-660.

Tsutsui, N., Saido-Sakanaka, H., Yang, W.J., Jayasankar, V., Jasmani, S., Okuno, A., Ohira, T., Okumura, T., Aida, K., and Wilder, M.N. (2004) Molecular characterization of a cDNA encoding vitellogenin in the coonstriped shrimp, *Pandalus hypsinotus* and site of vitellogenin mRNA expression. *J. Exp. Zoolog. A. Comp. Exp. Biol.* **301,** 802-814.

Tufail, M., Lee, J.M., Hatakeyama, M., Oishi, K., and Takeda, M. (2000) Cloning of vitellogenin cDNA of the American cockroach, *Periplaneta americana* (Dictyoptera), and its structural and expression analyses. *Arch. Insect Biochem. Physiol.* **45,** 37-46.

Van Antwerpen, R., Linnemans, W.A., Van der Horst, D.J., and Beenakkers, A.M.Th. (1988) Immunocytochemical localization of lipophorins in the flight muscles of the migratory locust (*Locusta migratoria*) at rest and during flight. *Cell Tissue Res.* **252,** 661-668.

Van Antwerpen, R. (1989) Negative staining of locust lipoproteins: implications for lipoprotein purification procedures. In: *Interaction of Locust Lipoproteins with Fat Body and the Flight Muscle.* Ph.D. Dissertation. Utrecht University, Utrecht, The Netherlands. 17–28.

Van der Horst, D.J. (1990) Lipid transport function of lipoproteins in flying insects. *Biochim. Biophys. Acta* **1047,** 195-211.

Van der Horst, D.J., Weers, P.M.M., and Van Marrewijk, W.J.A. (1993) Lipoproteins and lipid transport. In: *Insect Lipids: Chemistry, Biochemistry, and Biology.* (Stanley-Samuelson, D.W., and Nelson, D.R., Editors) University of Nebraska Press. 1–24.

Van der Horst, D.J., Van Marrewijk, W.J.A, and Diederen, J.H.B. (2001) Adipokinetic hormones of insect: release, signal transduction, and responses. *Int. Rev. Cytol.* **211,** 179-240.

Van der Horst, D.J., Van Hoof, D., Van Marrewijk, W.J.A., Rodenburg, K.W. (2002) Alternative lipid mobilization: the insect shuttle system. *Mol. Cell. Biochem.* **239,** 113-119.

Van der Horst, D.J., and Ryan, R.O. (2004) *Lipid transport.* In: *Comprehensive Molecular Insect Science* (Gilbert, L.I., Iatrou, K. and Gill, S.S., Editors) Elsevier. **4,** 225–246.

Van Heusden, M.C., Thompson, F., and Dennis, J. (1998) Biosynthesis of *Aedes aegypti* lipophorin and gene expression of its apolipoproteins. *Insect Biochem. Mol. Biol.* **28,** 733-738.

Van Hoof, D., Rodenburg, K.W., and Van der Horst, D.J. (2002) Insect lipoprotein follows a transferrin-like recycling pathway that is mediated by the insect LDL receptor homologue. *J. Cell Sci.* **115,** 4001–4012.

Van Hoof, D., Rodenburg, K.W., and Van der Horst, D.J. (2003) Lipophorin receptor-mediated lipoprotein endocytosis in insect fat body cells. *J. Lipid Res.* **44,** 1431-1440.

Van Hoof, D., Rodenburg, K.W., and Van der Horst, D.J. (2005a) Receptor-mediated endocytosis and intracellular trafficking of lipoproteins and transferrin in insect cells. Receptor-mediated endocytosis of insect lipoprotein. *Insect Biochem. Mol. Biol.* **35,** 117-128.

Van Hoof, D., Rodenburg, K.W., and Van der Horst, D.J. (2005b) Intracellular fate of LDL receptor family members depends on the cooperation between their ligand-binding and EGF domains. *J. Cell Sci.* **118,** 1309-1320.

Venkatesh, K., Lenz, C.J., Bergman, D.K., and Chippendale, G.M. (1987) Synthesis and release of lipophorin in larvaea of the southwestern corn borer, *Diatraea grandiosella*: an *in vitro* study. *Insect Biochem.* **17,** 1173-1180.

Wahli, W., Dawid, I.B., Wyler, T., Jaggi, R.B., Weber, R., and Ryffel, G.U. (1979) Vitellogenin in *Xenopus laevis* is encoded in a small family of genes. *Cell* **16,** 535-549.

Walzem, R.L., Hansen, R.J., Williams, D.L., and Hamilton, R.L. (1999) Estrogen induction of VLDLy assembly in egg-laying hens. *J. Nutr.* **129,** 467S-472S.

Wang, H., Tan, J.T., Emelyanov, A., Korzh, V., and Gong, Z. (2005) Hepatic and extrahepatic expression of vitellogenin genes in the zebrafish, *Danio rerio*. *Gene* **356,** 91-100.

Wang, L., Atkinson, D., and Small, D.M. (2003) Interfacial properties of an amphipathic alpha-helix consensus peptide of exchangeable apolipoproteins at air/water and oil/water interfaces. *J. Biol. Chem.* **278,** 37480-37491.

Wang, L., and Small, D.M. (2004) Interfacial properties of amphipathic β strand consensus peptides of apolipoprotein B at oil/water interfaces. *J. Lipid Res.* **45,** 1704-1715.

Weers, P.M.M., Van der Horst, D.J., Van Marrewijk, W.J.A., Van den Eijnden, M., Van Doorn, J.M., and Beenakkers, A.M.Th. (1992) Biosynthesis and secretion of insect lipoprotein. *J. Lipid Res.* **33,** 485-491.

Weers, P.M.M., Van Marrewijk, W.J.A., Beenakkers, A.M.Th., and Van der Horst, D.J. (1993) Biosynthesis of locust lipophorin. Apolipophorins I and II originate from a common precursor. *J. Biol. Chem.* **268,** 4300-4303.

Weers, P.M.M, and Ryan, R.O. (2003) Apolipophorin III: a lipid-triggered molecular switch. *Insect Biochem. Mol. Biol.* **33,** 1249-1260.

Wetterau, J.R., Combs, K.A., Spinner, S.N., and Joiner, B.J. (1990) Protein disulfide isomerase is a component of the microsomal triglyceride transfer protein complex. *J. Biol. Chem.* **265,** 9800-9807.

Wetterau, J.R., Combs, K.A., McLean, L.R., Spinner, S.N., and Aggerbeck, L.P. (1991) Protein disulfide isomerase appears necessary to maintain the catalytically active structure of the microsomal triglyceride transfer protein. *Biochemistry*. **30,** 9728-9735.

Willnow, T.E. (1999) The low-density lipoprotein receptor gene family: multiple roles in lipid metabolism. *J. Mol. Med.* **77,** 306-315.

Winter C.E., Penha C., and Blumenthal, T. (1996) Comparison of a vitellogenin gene between two distantly related rhabditid nematode species. *Mol. Biol. Evol.* **13,** 674-684.

Yano, K., Sakurai, M.T., Watabe, S., Izumi, S., and Tomino, S. (1994) Structure and expression of mRNA for vitellogenin in *Bombyx mori*. *Biochim. Biophys. Acta* **1218,** 1-10.

Yao, Z.M., Blackhart, B.D., Linton, M.F., Taylor, S.M., Young, S.G., and McCarthy, B.J. (1992) Expression of carboxyl-terminally truncated forms of human apolipoprotein B in rat hepatoma cells. Evidence that the length of apolipoprotein B has a major effect on the buoyant density of the secreted lipoproteins. *J. Biol. Chem.* **266,** 3300-3308.

Yeh, M.S., Huang, C.J., Leu, J.H., Lee, Y.C., and Tsai, I.H. (1999) Molecular cloning and characterization of a hemolymph clottable protein from tiger shrimp (*Penaeus monodon*). *Eur. J. Biochem.* **266,** 624-633.

Samenvatting

Elk organisme heeft vetten en andere lipiden nodig, aangezien deze componenten functioneren als bouwstenen van membranen, en als signaalmoleculen, energiebron en energiereserve. De slechte oplosbaarheid van lipiden in water bemoeilijkt echter hun transport. Dieren hebben een unieke oplossing gevonden om desondanks lipiden te kunnen distribueren via hun lichaamsvloeistof: lipoproteïnen.

Lipoproteïnen zijn partikels met een doorsnee tot circa 1 micrometer, die bestaan uit één of enkele eiwit-moleculen en vele lipid moleculen, in sommige gevallen honderden. Ze zijn vooral bestudeerd in zoogdieren in verband met hun rol in medische aandoeningen zoals atherosclerose, obesitas, en diabetes. Lipoproteïnen zijn echter ook onderzocht in andere, sterk verschillende dierklassen, met name in insecten. In deze oude en diverse diergroep is het functioneren van lipoproteïnen vooral bestudeerd in de context van langdurige vliegactiviteit, onder meer bij de treksprinkhaan *Locusta migratoria* tijdens diens massale trekvluchten.

Het (enige) lipoproteïne van insecten, genaamd lipoforine, beschikt in vergelijking met de vet-transporterende lipoproteïnen van zoogdieren over enkele opmerkelijke functionele en structurele eigenschappen. Zo is het herbruikbaar, waardoor lipid transport niet afhankelijk is van de synthese van nieuwe lipoproteïnen. Tevens verschillen de lipoproteïnen van insecten in hun componenten. Terwijl triacylglycerol het voornaamste lipid in zoogdier lipoproteïnen is, bevat lipoforine met name diacylglycerol, en bindt het ook beduidend minder lipid moleculen. Bovendien bevat lipoforine twee eiwitten, apoLp-I en apoLp-II, terwijl de vet-transporterende lipoproteïnen van zoogdieren gestabiliseerd worden door één eiwitcomponent, apoB. Het is echter gebleken dat apoLp-I en -II ontstaan uit de klieving van een precursor eiwit, genaamd apoLp-II/I, dat evolutionair verwant is aan apoB. ApoLp-II/I en apoB maken deel uit van de familie van ‘large lipid transfer’ eiwitten, de LLTP, die gekarakteriseerd worden door homologie in de N-terminale ~900 aminozuren van deze eiwitten, de LLT module.

De specifieke eigenschappen van insect lipoproteïne ten opzichte van de vet-transporterende zoogdier lipoproteïnen vinden mogelijk hun oorsprong in verschillen tussen de structuren waaruit ze zijn opgebouwd, en vanuit dit perspectief zijn de biosynthese van lipoforine, de evolutie van lipoproteïne precursors, alsmede de diversiteit in hun receptoren onderzocht in dit proefschrift.

Lipoforine biosynthese heeft twee opvallende kenmerken: 1) klieving van apoLp-II/I in apoLp-I en –II, en 2) lipidatie van apoLp-II/I of de klievingsproducten daarvan tot een lipoproteïne. Deze processen vinden plaats in het vetlichaam, een weefsel dat functies van de lever en het vetweefsel van zoogdieren (en andere gewervelde dieren) combineert.

In Hoofdstuk 2 wordt de klieving van apoLp-II/I in apoLp-I en apoLp-II nader bestudeerd. Hiertoe werd een recombinant insecten systeem opgezet dat een N-terminaal deel van apoLp-II/I, overeenkomend met het volledige apoLp-II en een deel van apoLp-I, tot expressie brengt. Door specifieke aminozuren in apoLp-II/I te wijzigen kon klieving worden voorkomen, evenals door incubatie van recombinante cellen met een remmer voor proproteïne convertases (een groep van eiwit-klievende enzymen). Deze experimenten ondersteunen de betrokkenheid van een furine-achtige proproteïne convertase bij de klieving van apoLp-II/I. *In vitro* incubatie van de klievingsremmer met *L. migratoria* vetlichaam wees uit dat klieving niet essentieel is voor de secretie en lipidatie van lipoforine, aangezien ongekliefd apoLp-II/I tevens een hoge-dichtheids lipoforine blijkt te vormen. Deze vinding roept de vraag op wat dan de functie van apoLp-II/I klieving is.

Modellering van de LLT module van apoLp-II/I, inclusief apoLp-II en een deel van apoLp-I, gaf structurele gelijkenis aan met apoB, het structurele eiwit van vele zoogdier lipoproteïnen (Hoofdstuk 2). Uitgaande van kennis over apoB lipoproteïne biosynthese werd de lipidatie van apoLp-II/I nader onderzocht in Hoofdstuk 3. Amfipatische structuren spelen een zeer belangrijke rol in de associatie van lipid, zowel in apoB als in vele andere ongerelateerde lipid-bindende eiwitten. Voorspelling van de structuur wijst op de aanwezigheid van clusters verrijkt in amfipatische α-helices danwel amfipatische β-strands in apoLp-II/I. Deze zijn, bezien vanaf de N- tot de C-terminus, georganiseerd als α-β-α, enigszins gelijkend op een verkorte versie van apoB, dat de organisatie α-β-α-β-α bezit. Recombinant expressie van truncaties van apoLp-II/I wijst uit dat de sequentie van het β cluster noodzakelijk is voor lipidatie tot een lipoproteïne.

De factor MTP (microsomaal triglyceride transfer proteïne) speelt een essentiële rol in de lipidatie van apoB tot een lipoproteïne. Aangezien insecten ook een MTP gen bezitten, werd de mogelijke rol daarvan in lipoforine biosynthese onderzocht (Hoofdstuk 3). Co-expressie van een insecten MTP blijkt de lipoproteïne vorming te stimuleren. Niet alleen de secretie van apoLp-II/I en diens klievingsproducten bleek toe te nemen, maar ook hun lipidatie. Hoewel het belang van MTP in lipoproteïne biosynthese nog vastgesteld dient te worden *in vivo*, wijzen deze vinding én de rol van het amfipatische β cluster in lipid binding er sterk op dat de biosynthese van lipoproteïnen in zowel insecten als zoogdieren afhankelijk is van dezelfde structurele componenten. De huidige mechanismen van lipoproteïne biogenese vinden dus hun oorsprong in de eerste dieren.

Om inzicht te verwerven in de origine en de huidige variatie van lipoproteïnen in dieren werd de evolutie en diversiteit van de familie waartoe apoB en apoLp-II/I behoren, de

LLTP familie, nader bestudeerd (Hoofdstuk 4). Hoewel de oorsprong van deze familie niet getraceerd kon worden, kwam uit fylogenetische analyse naar voren dat de LLTP familie opgedeeld kan worden in 3 subfamilies, te weten de apoB-achtige LLTP, de vitellogenine-achtige LLTP, én de MTP's. ApoLp-II/I en apoB behoren tot de apoB-achtige LLTP, evenals een nieuw geïdentificeerd LLTP in insecten, voorlopig het 'apoLp-II/I-related protein' genoemd. Opvallend genoeg behoort het voornaamste dooiereiwit in eicellen van een groep van kreeftachtigen, dat daar vitellogenine was genoemd, ook tot de apoB-achtige LLTP. De apoB-achtige LLTP vertonen unieke gelijkenissen in de LLT module alsmede in een geconserveerd motief daar juist buiten, en in de relatieve grootte van hun amfipatische clusters. De functionele en structurele diversiteit van LLTP wordt bediscussieerd vanuit een evolutionair perspectief.

Waar de LLTP de structurele basis vormen voor lipoproteïnen en hun biogenese, wordt de cellulaire opname van lipoproteïnen gemedieerd door leden van de LDLR familie. De huidige classificatie van nieuw-ontdekte LDLR familieleden is vaak gebaseerd op het aantal cysteïne-rijke eenheden in het ligand-bindend domein. In Hoofdstuk 5 laten we zien dat ze ook geclassificeerd kunnen worden (en wellicht beter) op basis van hun C-terminale niet-repetitieve domeinen. Op deze wijze worden insect lipoforine receptoren, die onderling verschillen in het aantal cysteïne-rijke eenheden in het ligand-bindend domein, gekarakteriseerd als een aparte groep. De unieke sequentiekarakteristieken van LpR's in deze domeinen worden besproken in relatie tot hun functionele eigenschappen ten opzichte van de LDLR familieleden in zoogdieren.

Samengevat zijn de belangrijkste resultaten beschreven in dit proefschrift: (1) het structurele eiwit van lipoforine, apoLp-II/I, wordt gekliefd door een insecten furine, (2) klieving is niet essentieel voor de vorming van apoLp-II/I tot een hoge-dichtheids lipoproteïne, (3) lipoproteïne vorming is afhankelijk van de aanwezigheid van een regio die voorspeld werd verrijkt te zijn in amfipatische β-strands, (4) MTP stimuleert de secretie en lipidatie van lipoforine in een recombinant expressie systeem, (5) apoLp-II/I is een apoB-achtige LLTP, en (6) LDLR familie leden kunnen worden geklassificeerd op basis van hun niet-repetitieve domeinen.

Aldus toont dit proefschrift dat de biogenese van lipoproteïnen en het lipid transport in zowel zoogdieren als insecten geregeerd wordt door gemeenschappelijke moleculaire structuren en mechanismen. Modificatie van deze elementen, door een samenspel van evolutie en biosynthese, resulteert in de compositionele en functionele verschillen tussen lipoproteïnen. Vanuit dit inzicht alsmede de ontwikkelde methode tot recombinant lipoforine expressie kan nu worden voortgegaan naar het begrijpen van de moleculaire achtergrond van de opmerkelijke eigenschappen van de lipoproteïnen en het lipid transport in insecten.

Dankwoord

Zo, op naar de volgende uitdaging! Ik schrijf het ook tot mijn eigen verrassing, maar gelukkig voelt het echt zo. Het waren vier intensieve jaren, waarin velen hebben bijgedragen aan het tot stand komen van dit proefschrift. Bij deze wil ik dan ook iedereen bedanken die me geholpen heeft in dit onderzoek.

Mijn eerste dank gaat uit naar mijn promotor, Dick van der Horst. Dick, ik wil je danken voor de benodigde steun, goede atmosfeer, beslissingsruimte en deadlines. Met genoegen denk ik terug aan onze discussies over de vergelijkende en integratieve fysiologie. Mijn co-promotor, Kees Rodenburg, ik dank je voor de ruimte die je me liet, de stimulatie tot doorzetten, en voor het aanwakkeren van de biochemische invalshoek. Als begeleiders hebben jullie elkaar uitstekend gecomplementeerd, en ik ben jullie zeer erkentelijk dat ik ook mijn eigen draai aan dit proefschrift heb kunnen geven.

Vele andere leden van onze leerstoelgroep Stofwisselingsfysiologie hebben bijgedragen. Bovenal dank ik Marcelle Kasperaitis en Jana Kerver voor hun essentiële bijdragen aan de ontwikkeling van het recombinant lipoproteïne expressie systeem. Jan van Doorn bedank ik voor zijn behulpzame en bovenal wijze ondersteuning, ik zal je goedmoedige gegrom missen. Vele andere collega's hebben bijgedragen met nuttige feedback. Mijn oud-kamergenote Karine Valentijn dank ik specifiek voor haar nuttige kritische blik, en uiteraard het genoegen om een connaisseur van Seinfeld te treffen. Mijn latere kamergenoot, Koert Burger, bedank ik voor het meedenken en onze bijzonder aangename discussies; het zou mooi zijn om daar nog een experimenteel vervolg aan te kunnen geven. Wil van Marrewijk, je enthousiasme voor en kennis van onderwijs waren hartverwarmend. Dennis van Hoof, je bewonderenswaardige doelgerichtheid was inspirerend en leerzaam. En mede-promovendi Masja van Oort en Sigrid Roosendaal, ik zal het maar bekennen: na jaren van ‘verzet’ heb ik nu ook een agenda in gebruik genomen. Samen met Yvonne Derks hebben jullie de buurkamer aardig verlevendigd en jullie vrolijke klanken hebben me uit menige afterlunch-dip geschud, bij deze alsnog dank daarvoor.

Ole Madsen (Protein Biochemistry, Radboud Universiteit Nijmegen) en Paul Richardson (Department of Chemistry, Coastal Carolina University, USA) dank ik voor hun essentiële bijdrage aan respectievelijk de fylogenetische analyse en de modellering. Ole, bedankt voor je uitstekende analyses en prettige samenwerking. Je last-minute ondersteuning was geweldig, en ik hoop je ooit een wederdienst te kunnen bewijzen, naast

het co-auteurschap. Paul, our collaboration was very fruitful, even bringing our work on the cover of *Journal of Lipid Research*. All the best in your career!

Hans van Aken van de DNA sequencing faciliteit dank ik voor de vele met zorg uitgevoerde sequence runs. De dierverzorgers Henk Schriek, Ko van Rootselaar, Job Jansen† en later Cor Nijhoff dank ik voor de altijd levendige sprinkhanen. Ook Joke Granneman, Wouter Nijenhuis en Jan Bogerd van de groep Endocrinologie dank ik voor hun hulpvaardigheid als ik even iets moest 'lenen'.

Met veel plezier heb ik studenten Biologie en Biomedische Wetenschappen mogen begeleiden met practica en scripties, en hun inzet heeft zeer zeker bijgedragen tot dit eindresultaat. Vooral mijn master-student Antoine de Morrée heeft een grote bijdrage geleverd aan dit werk met zijn inzet, discussie, en belangstelling. Antoine, ik ben je erg dankbaar dat je me na afloop van je studie zelfs vrijwillig geholpen hebt, experimenteel, en bij de spellingscontrole op het eind. Ik kijk uit naar ons volgende diepgaande bargesprek.

Naast bovengenoemden was dit proefschrift echter nooit gereed gekomen zonder de aanwezigheid van vrienden en familie. Allemaal bedankt voor jullie belangstelling gedurende de gehele tijd! Ik hoop de komende periode wat meer tijd met jullie te kunnen besteden. Klaas, Arjen, Bas, Wouter en Geert, we moeten maar weer eens het immer dynamische vak Z bezoeken. Ward, Richard, Rob en Chantal, ik hoop jullie voortaan vaker te kunnen treffen in Heitse en contreien, of hier in Utrecht natuurlijk. Marjolijn, bedankt voor de onvergetelijke momenten, en uiteraard voor het accurate fractioneren van de nodige KBr-gradiënten. Jayand, bedankt voor de gezellige salsa-avondjes. Drie verhuizingen in vier jaar hebben me op allerlei plekken in Utrecht gebracht, en ik dank mijn oud-huisgenoten Marc en Coert voor alle gein buiten werktijd. János, je inspirerende vriendschap is me bijzonder dierbaar, egészségedre! En Roland en Martijn, mijn Limburgs-Utrechtse vrienden, ik vind het geweldig dat we zoveel met elkaar optrekken en elkaar aansporen in werkelijk alles. Ten slotte, maar eigenlijk op de eerste plaats, wil ik mijn ouders, en Erwin en Danny bedanken voor hun niet-aflatende steun en interesse. Jullie staan aan de basis van alles.

En nu op naar een nieuwe uitdaging, spannend!

Publications

Papers

Van Rheede, T., Smolenaars, M.M.W., Madsen, O., and De Jong, W.W. (2003) Molecular evolution of the mammalian prion protein. *Mol. Biol. Evol.* **20,** 111-121.

Mancera, J.M., Smolenaars, M., Laiz-Carrion R., Martin del Rio, M., Bonga, S.E.W., and Flik, G. (2004) 17β-Estradiol affects osmoregulation in *Fundulus heteroclitus*. *Comp. Biochem. Physiol. B* **139,** 183-191.

Smolenaars, M.M.W., Kasperaitis, M.A.M., Richardson, P.E., Rodenburg, K.W., and Van der Horst, D.J. (2005) Biosynthesis and secretion of insect lipoprotein: involvement of furin in cleavage of the apoB homolog, apolipophorin-II/I. *J. Lipid Res.* **46,** 412-421.

Rodenburg, K.W., Smolenaars, M.M.W., Van Hoof, D., and Van der Horst, D.J. (2006) Sequence analysis of the non-recurring C-terminal domains shows that insect lipoprotein receptors constitute a distinct group of LDL receptor family members. *Insect Biochem. Mol. Biol. In press.*

Smolenaars, M.M.W., de Morrée, A., Kerver, J., Rodenburg, K.W., and Van der Horst, D.J. (2006) Insect lipoprotein biosynthesis depends on microsomal triglyceride transfer protein and an amphipathic β cluster in apolipophorin-II/I. *To be submitted.*

Smolenaars, M.M.W., Madsen, O., Rodenburg, K.W., and Van der Horst, D.J. (2006) Molecular evolution and diversity of the large lipid transfer protein family. *To be submitted.*

Lectures

Smolenaars, M.M.W., Kasperaitis, M.A.M., Rodenburg, K.W. and Van der Horst D.J.: Lipoprotein assembly in insects vs. mammals: on the function of insect apolipophorin cleavage. International Symposium on Animal Physiology: Proteins in Adaptation and Evolution, Greifswald, Germany, June 3-5, 2004

Smolenaars, M.M.W.: Biosynthesis of lipoprotein particles in insects: homologies and adaptations. Study Group on Lipids and Biomembranes, Lunteren, March 14-15, 2005.

Smolenaars, M.M.W.: Evolution of lipoproteins: homologies and adaptations. 12th Benelux Congress of Zoology, Wageningen, October 28, 2005.

Curriculum vitae

Marcel Smolenaars werd op 20 april 1978 in Weert geboren. Na het atheneum afgerond te hebben aan de scholengemeenschap St.-Ursula te Horn, is hij in 1996 begonnen aan de studie Biologie aan de Radboud Universiteit te Nijmegen. Tijdens deze studie werden verschillende stages uitgevoerd. Bij de afdeling Organismale Dierfysiologie werd gewerkt aan de moleculaire karakterisatie van een stress hormoon in de karper, en bij de afdeling Eiwit Biochemie aan de moleculaire evolutie van prionen in zoogdieren. Tevens werd een stage volbracht aan de Universidad de Cádiz (Spanje) aan osmoregulatie in vissen. Na het behalen van zijn doctoraal examen in 2001 is hij werkzaam geweest als promovendus bij de leerstoelgroep Stofwisselingsfysiologie van het departement Biologie van de Universiteit Utrecht. De resultaten beschreven in dit proefschrift zijn in deze periode verkregen. De auteur was gedurende deze periode lid van het Instituut voor Biomembranen.

Marcel Smolenaars was born on 20 April 1978 in Weert, The Netherlands. After finishing his secondary education, he started in 1996 with the study Biology at the University of Nijmegen, The Netherlands. As an undergraduate student he participated in several studies. At the department of Organismal Animal Physiology he participated in the molecular characterization of a stress hormone in carp, and at the department of Protein Biochemistry he studied the molecular evolution of prion proteins in mammals. In addition, he studied osmoregulation in fish at the University of Cádiz (Spain). Following graduation in 2001, he enrolled as a Ph.D. student at the group of Biochemical Physiology at the department of Biology of Utrecht University. The results of his studies during this period are described in this thesis. During this period he was a member of the Institute of Biomembranes.