

Review

## Comparative Studies of Vertebrate Platelet Glycoprotein 4 (CD36)

Roger S. Holmes

Eskitis Institute for Cell and Molecular Therapies, Griffith University, Nathan, QLD 4111, Australia;  
E-Mail: r.holmes@griffith.edu.au; Tel.: +61-7-3735-7773

Received: 2 August 2012; in revised form: 6 September 2012 / Accepted: 18 September 2012 /

Published: 24 September 2012

---

**Abstract:** Platelet glycoprotein 4 (CD36) (or fatty acyl translocase [FAT], or scavenger receptor class B, member 3 [SCARB3]) is an essential cell surface and skeletal muscle outer mitochondrial membrane glycoprotein involved in multiple functions in the body. CD36 serves as a ligand receptor of thrombospondin, long chain fatty acids, oxidized low density lipoproteins (LDLs) and malaria-infected erythrocytes. CD36 also influences various diseases, including angiogenesis, thrombosis, atherosclerosis, malaria, diabetes, steatosis, dementia and obesity. Genetic deficiency of this protein results in significant changes in fatty acid and oxidized lipid uptake. Comparative CD36 amino acid sequences and structures and *CD36* gene locations were examined using data from several vertebrate genome projects. Vertebrate CD36 sequences shared 53–100% identity as compared with 29–32% sequence identities with other CD36-like superfamily members, *SCARB1* and *SCARB2*. At least eight vertebrate CD36 *N*-glycosylation sites were conserved which are required for membrane integration. Sequence alignments, key amino acid residues and predicted secondary structures were also studied. Three CD36 domains were identified including cytoplasmic, transmembrane and exoplasmic sequences. Conserved sequences included *N*- and *C*-terminal transmembrane glycines; and exoplasmic cysteine disulphide residues; TSP-1 and PE binding sites, Thr92 and His242, respectively; 17 conserved proline and 14 glycine residues, which may participate in forming CD36 ‘short loops’; and basic amino acid residues, and may contribute to fatty acid and thrombospondin binding. Vertebrate *CD36* genes usually contained 12 coding exons. The human *CD36* gene contained transcription factor binding sites (including PPAR $\gamma$  and PPAR $\alpha$ ) contributing to a high gene expression level (6.6 times average). Phylogenetic analyses examined the relationships and potential evolutionary origins of the vertebrate *CD36* gene with vertebrate *SCARB1* and *SCARB2* genes. These suggested that *CD36* originated in an ancestral genome and was subsequently duplicated to form three vertebrate *CD36* gene family members, *SCARB1*, *SCARB2* and *CD36*.

**Keywords:** vertebrates; amino acid sequence; CD36; evolution; thrombospondin receptor

---

## 1. Introduction

Platelet glycoprotein 4 (CD36) (cluster of differentiation 36) (or fatty acyl translocase [FAT]; and scavenger receptor class B, member 3 [SCARB3]) is one of at least three members of the CD36-like family that is an integral membrane protein of many tissues of the body which plays a role in fatty acyl translocation and as a multiple ligand cell surface receptor of oxidized LDL lipoproteins (ox-LDL), long chain fatty acids, aged neutrophils and *Plasmodium falciparum*-parasitized erythrocytes (PE) which has been implicated in several diseases including insulin resistance, diabetes, atherosclerosis and malaria [1–10]. CD36 has also been reported on the outer mitochondrial membrane of skeletal muscle and serves a long chain fatty acid transport role, as well as contributing to the regulation of fatty acid oxidation by muscle mitochondria [11]. In addition, CD36 contributes to cerebrovascular oxidative stress and neurovascular dysfunction induced by amyloid-beta in Alzheimer's dementia [12,13] and may serve a 'lipid-sensing' role in the body with a broad physiological role as a lipid-receptor protein which influences eating behavior and energy balance [14]. Moreover, a specific CD36-dependent signaling pathway has been proposed for platelet activation by ox-LDL [15].

SCARB1 (also called CLA1, SRB1 and CD36L1), a second member of the CD36-like family, is a homo-oligomeric plasma membrane cell surface glycoprotein receptor for high density lipoprotein cholesterol (HDL), other phospholipid ligands and chylomicron remnants [16–20]. SCARB2 (also called LIMP2 (lysosomal integral membrane protein), SRB2 and CD36L2) is a third member of the CD36 family predominantly integrated within lysosomal and endosomal membranes which contributes to lysosomal membrane organization and transport functions [21–25].

The gene encoding CD36 (*CD36* in humans; *Cd36* in mice) is localized on chromosome 7 q11.2 and is encoded by 15 exons, including 12 coding exons [26–29]. Human *CD36* is expressed at very high levels in various cells and tissues of the body, including platelets, monocytes/macrophages, and microvascular endothelial cells, plays important roles in atherosclerosis, inflammation, thrombosis and angiogenesis [4,6,7,30–32], and is upregulated in human monocytes following statin administration [33]. Studies of *Cd36*<sup>-/-</sup>/*Cd36*<sup>-/-</sup> knockout mice have shown that *CD36*-deficiency protects against Western-type diet related cardiac dysfunction [34–36] and contributes to a reduction in fatty acid oxidation by muscle mitochondria [11,37]. Human clinical studies have also examined *CD36* polymorphisms associated with enhanced atherosclerotic cardiovascular diseases [38,39], type II diabetes [9], oral fat perception, fat preference and obesity in African-Americans [40] and protection from malaria [41,42]. In addition, hepatic *CD36* upregulation has been shown to be associated with insulin resistance, hyperinsulinaemia, and increased steatosis in patients with non-alcoholic steatohepatitis and chronic hepatitis C [43]. Reviews of the role of macrophage human CD36 in atherosclerosis have been published [7,44].

This paper reports the predicted gene structures and amino acid sequences for several vertebrate *CD36* genes and proteins, the secondary structures for vertebrate CD36 proteins, several potential sites for regulating human *CD36* gene expression and the structural, phylogenetic and evolutionary

relationships for these genes and enzymes with those for vertebrate *CD36*, *SCARB1* and *SCARB2* gene families.

## 2. Results and Discussion

### 2.1. Alignments of Vertebrate *CD36* Amino Acid Sequences

The deduced amino acid sequences for cow (*Bos taurus*), opossum (*Monodelphis domestica*), chicken (*Gallus gallus*), frog (*Xenopus tropicalis*) and zebrafish (*Danio rerio*) *CD36* are shown in Figure 1 together with previously reported sequences for human and mouse *CD36* (Table 1) [45,46]. Alignments of human with other vertebrate *CD36* sequences examined were 53–100% identical, suggesting that these are products of the same family of genes, whereas comparisons of sequence identities of vertebrate *CD36* proteins with human *SCARB1* and *SCARB2* proteins exhibited lower levels of sequence identities (30–32%), indicating that these are members of distinct *CD36*-like gene families (Supplementary Table 1).

The amino acid sequences for eutherian mammalian *CD36* contained 472 residues, whereas opossum (*Monodelphis domestica*), platypus (*Ornithorhynchus anatinus*) and chicken (*Gallus gallus*) *CD36* sequences contained 471 residues, while frog (*Xenopus tropicalis*) and zebrafish (*Danio rerio*) *CD36* sequences contained 470 and 465 amino acids, respectively (Table 1; Figure 1). Previous studies have reported several key regions and residues for human and mouse *CD36* proteins (human *CD36* amino acid residues were identified in each case). These included cytoplasmic *N*-terminal and *C*-terminal residues: residues 2-6 and 462-472; *N*-terminal and *C*-terminal trans-membrane helical regions: residues 7-28 and 440-461 [32,45]; palmitoylated cysteine residues (Cys3; Cys7; Cys464; and Cys466) in the *N*- and *C*-terminal *CD36* cytoplasmic tails [47]; exoplasmic Thr92, which is phosphorylated by protein kinase C alpha and contributes to the suppression of thrombospondin-1 binding *in vitro* [48]; His242 which contributes to the interaction of *CD36*-dependent endothelial cell adherence with *Plasmodium falciparum* [4]; and six exoplasmic disulfide bond forming residues: Cys243, Cys272, Cys311, Cys313, Cys322 and Cys333 [49].

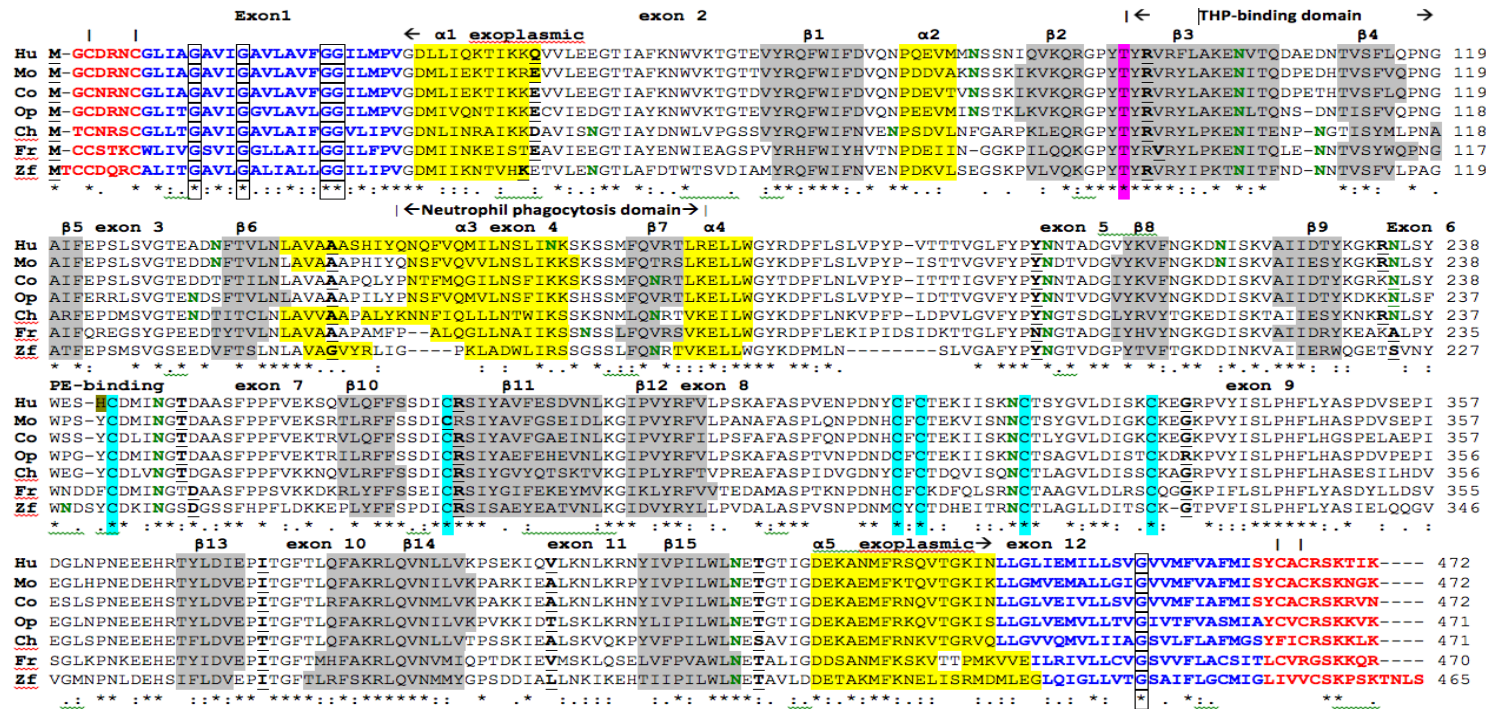
**Table 1.** *CD36*, *SCARB1* and *SCARB2* genes and proteins. RefSeq: the reference amino acid sequence; Ꞥredicted Ensembl amino acid sequence; na-not available; GenBank IDs are derived from NCBI <http://www.ncbi.nlm.nih.gov/genbank/>; Ensembl ID was derived from Ensembl genome database <http://www.ensembl.org>; \* designates scaffold; Un refers to unknown chromosome; UNIPROT refers to UniprotKB/Swiss-Prot IDs for individual CD36-like proteins (see <http://kr.expasy.org>); Un-refers to unknown chromosome; bps refers to base pairs of nucleotide sequences; the number of coding exons are listed; gene expression levels are in **bold**.

<i>CD36 Gene</i>	Species	RefSeq ID Ensembl/NCBI	GenBank ID	UNIPROT ID	Amino acids	Chromosome location	Coding Exons (strand)	Gene Size bps	Gene Expression Level
<b>Human</b>	<i>Homo sapiens</i>	NM_001001547	BC008406	P16671	472	7:80,275,645-80,303,732	12 (+ve)	72,231	<b>6.6</b>
<b>Chimpanzee</b>	<i>Pan troglodytes</i>	XP_519573	na	na	472	7:81,142,402-81,169,764	12 (+ve)	#27,363	na
<b>Orangutan</b>	<i>Pongo abelii</i>	XP_002818343	na	na	472	7:95,750,733-95,779,630	12 (-ve)	#28,898	na
<b>Gibbon</b>	<i>Nomascus leucogenys</i>	XP_003252221	na	na	472	*GL397261:11,570,433-11,598,114	12 (+ve)	#27,682	na
<b>Rhesus</b>	<i>Macaca mulatta</i>	NP_001028085	na	na	472	3:136,626,102-136,653,066	12 (+ve)	#27,682	na
<b>Mouse</b>	<i>Mus musculus</i>	NM_001159555.1	BC010262	Q08857	472	5:17,291,543-17,334,712	12 (-ve)	43,170	<b>4.2</b>
<b>Rat</b>	<i>Rattus norvegicus</i>	NP_113749	L19658	Q07969	472	4:13,472,534-13,522,334	12 (+ve)	49,801	<b>0.3</b>
<b>Guinea Pig</b>	<i>Cavia porcellus</i>	XP_003469862	na	na	472	*31:20,074,611-20,098,210	12 (+ve)	#23,600	na
<b>Cow</b>	<i>Bos taurus</i>	NM_17410	BC103112	P26201	472	4:40,585,624-40,614,621	12 (-ve)	#28,998	na
<b>Dog</b>	<i>Canis familiaris</i>	NM_001177734	ADE58431	na	472	18:23,334,171-23,360,045	12 (+ve)	#25,875	na
<b>Pig</b>	<i>Sus scrofa</i>	NP_001038087	AK400585	Q3HUX1	472	9:93,204,848-93,241,842	12 (-ve)	#36,995	na
<b>Rabbit</b>	<i>Oryctolagus cuniculus</i>	XP_002712062	na	na	472	7:35,303,111-35,333,630	12 (-ve)	#30,520	na
<b>Horse</b>	<i>Equus caballus</i>	XP_001487957	na	na	472	4:6730,96-698,607	12 (-ve)	#25,512	na
<b>Elephant</b>	<i>Loxodonta africana</i>	XP_003407226	na	na	472	5: 69,036,730-69,073,879	12 (-ve)	#37,150	na
<b>Opossum</b>	<i>Monodelphis domestica</i>	XP_001364375	na	na	471	8:149,041,138-149,075,533	12 (-ve)	#34,396	na
<b>Platypus</b>	<i>Ornithorhynchus anatinus</i>	XP_001506583	na	na	471	*Ultra5:3,505,963-3,536,963	12 (-ve)	#31,001	na
<b>Chicken</b>	<i>Gallus gallus</i>	ENSGALG8439	AJ719746	F1NER9	471	1:12,077,308-12,107,415	12 (-ve)	30,108	na
<b>Lizard</b>	<i>Anolis carolinensis</i>	XP_003221568	na	na	472	5:93,087,943-93,120,933	12 (-ve)	#32,991	na
<b>Frog</b>	<i>Xenopus tropicalis</i>	NP_001107151	na	na	470	*GL172681:663,550-679,762	12 (-ve)	#16,213	na
<b>Zebrafish</b>	<i>Danio rerio</i>	NP_001002363.1	BC076048	Q6DHC7	465	4:21,594,449-21,606,961	12 (-ve)	12,513	na

Table 1. Cont.

<b>SCARB1 Gene</b>	<b>Species</b>	<b>RefSeq ID Ensembl/NCBI</b>	<b>GenBank ID</b>	<b>UNIPROT ID</b>	<b>Amino acids</b>	<b>Chromosome location</b>	<b>Coding Exons (strand)</b>	<b>Gene Size bps</b>	<b>Gene Expression Level</b>
<b>Human</b>	<i>Homo sapiens</i>	NM_00505	BC022087	Q8WVT0	509	12:125,267,232-125,348,266	12 (-ve)	81,035	<b>13.7</b>
<b>Mouse</b>	<i>Mus musculus</i>	NM_001205082.1	BC004656	Q61009	509	5:125,761,478-125,821,252	12 (-ve)	63,985	<b>5.1</b>
<b>Chicken</b>	<i>Gallus gallus</i>	XP_415106	na	na	503	15:4,543,054-4,558,954	12 (+ve)	15,901	na
<b>Zebrafish</b>	<i>Danio rerio</i>	NM_198121	BC044516	E7FB50	496	11:21,526,513-21,572,478	12 (-ve)	45,684	na
<b>SCARB2 Gene</b>									
<b>Human</b>	<i>Homo sapiens</i>	NM_005506	BT006939	Q53Y63	478	4:77,084,378-77,134,696	12 (-ve)	50,316	<b>3.2</b>
<b>Mouse</b>	<i>Mus musculus</i>	NM_007644	BC029073	O35114	478	5:92,875,330-92,934,334	12 (-ve)	59,005	<b>3.6</b>
<b>Chicken</b>	<i>Gallus gallus</i>	XP_42093.1	BX931548	na	481	4:51,411,268-51,429,620	12 (+ve)	18,353	na
<b>Zebrafish</b>	<i>Danio rerio</i>	NM_173259.1	BC162407	Q8JQR8	531	5: 63,942,096-63,955,449	13 (+ve)	13,354	na
<b>CD36 Gene</b>									
<b>Lancelet</b>	<i>Branchiostoma floridae</i>	XP_002609178.1	na	na	480	Un:534,334,234-534,343,082	12 (+ve)	8,849	na
<b>Sea squirt</b>	<i>Ciona intestinalis</i>	XP_002127015.1	na	na	523	09p:2,872,362-2,873,903	1 (-ve)	1,542	na
<b>Nematode</b>	<i>Caenorhabditis elegans</i>	NM_067224	na	Q9XTT3	534	III:12,453,609-12,456,726	8 (+ve)	3,118	<b>4.6</b>
<b>Fruit fly</b>	<i>Drosophila melanogaster</i>	NP_523859	na	na	520	2R:20,864,606-20,867,116	6 (-ve)	#2,511	na

**Figure 1.** Amino Acid Sequence Alignments for Vertebrate CD36 Sequences. See Table 1 for sources of CD36 sequences; Hu-human; Mo-mouse; Co-cow; Op-opossum; Ch-chicken; Fr-frog; Zf-zebrafish; \* shows identical residues for CD36 subunits; : similar alternate residues; . dissimilar alternate residues; predicted cytoplasmic residues are shown in red; predicted transmembrane residues are shown in blue; N-glycosylated and potential N-glycosylated Asn sites are in green; exoplasmic Thr92, which is phosphorylated by pyruvate kinase alpha, is shown in pink; predicted disulfide bond Cys residues are shown in blue; predicted  $\alpha$ -helices for vertebrate CD36 are in shaded yellow and numbered in sequence from the start of the predicted exoplasmic domain; predicted  $\beta$ -sheets are in shaded grey and also numbered in sequence; bold underlined font shows residues corresponding to known or predicted exon start sites; exon numbers refer to human CD36 gene exons; G residues refer to conserved glycines in the N- and C-terminal oligomerisation domains of the trans-membrane sequence [49]; CD36 binding domains are identified: THP-refers to binding region for low-density lipoproteins [6–8]; neutrophil phagocytosis domain designated by [3,7]; PE binding refers to cytoadherence region of *Plasmodium falciparum*-parasitized erythrocytes (PE) to endothelial cells [4].



## 2.2. Comparative Sequences for Vertebrate CD36 N-Glycosylation Sites

Ten exoplasmic *N*-glycosylation sites for human CD36 have been previously identified for this protein (Figure 1; Table 2) [50]. One of these sites (site 2) contained a proline residue at the second position and may not function as an *N*-glycosylation site due to proline-induced inaccessibility [51]. Eight of these sites were predominantly retained among the 19 vertebrate CD36 sequences examined (sites 4, 5, 10, 15, 19, 23 and 25) (Figure 1; Table 2). The sequence conservation observed for these residues among the vertebrate CD36 sequences examined suggests that they contribute significantly to the structure and function of vertebrate CD36 as a glycoprotein. The multiple *N*-glycosylation sites observed for vertebrate CD36 sequences suggest a role for *N*-proteoglycan residues exposed on the external surface of plasma membranes in the performance of CD36 functions in binding various lipid molecules, including long chain fatty acids. This is also supported by recent animal model studies examining the impacts of reduced *N*-glycosylation upon cardiac long chain fatty metabolism, which demonstrated a key role for *N*-glycosylation in the recruitment of CD36 into cardiac membranes [52].

## 2.3. Conserved Glycines in the N-Terminal Domain of the CD36 Trans-Membrane Sequence

The *N*-terminal region for vertebrate CD36 sequences (residues 1-29 for human CD36) contained cytoplasmic (residues 2-7) and trans-membrane (residues 8-29) motifs which underwent changes in amino acid sequence but retained predicted cytoplasmic and trans-membrane properties in each case, respectively (Figure 1). Vertebrate *N*-terminal trans-membrane sequences, in particular, were predominantly conserved, especially for CD36 Gly12, Gly16 and Gly24/Gly25 residues, which were observed among the vertebrate CD36 sequences examined (Figure 1). Site directed mutagenesis studies of the related human SCARB1 sequence have demonstrated key roles for *N*-terminus trans-membrane sequence glycine residues, by facilitating oligomerisation and selective lipid uptake by SCARB1 conserved glycine residues [53] and similar roles may apply to the conserved *C*-terminal domain CD36 glycine residues. A recent report has shown, however, that CD36 is capable of binding acetylated and oxidized low-density lipoproteins as a monomer, even though multiple homo- and hetero-protein interactions are formed in the plasma membrane [8]. A conserved glycine residue was also observed for the vertebrate *C*-terminal trans-membrane sequences (human CD36 Gly452) (Figure 1), however the role for this residue has not been investigated.

**Table 2.** Predicted *N*-glycosylation sites for CD36 sequences. Numbers refer to amino acids in the acid sequences, including *N*-asparagine; K-lysine; I-isoleucine; H-histidine; S-serine; T-threonine; Q-glutamine; D-aspartate; Y-tyrosine; and V-valine. Note that there are 25 potential sites identified for vertebrate CD36 and other CD36-like sequences, including 10 sites for human CD36 (see [49]). *N*-glycosylation sites were identified using the NetNGlyc 1.0 web server (<http://www.cbs.dtu.dk/services/NetNGlyc/>). Higher probability *N*-glycosylation sites are in **bold**.

Vertebrate	Species	Site 1	Site 2*	Site 3	Site 4	Site 5	Site 6	Site 7	Site 8	Site 9	Site 10	Site 11	Site12	Site 13
CD36														
Human	<i>Homo sapiens</i>				<b>79N</b> SSN	<b>102N</b> V <b>TQ</b>					<b>134N</b> F <b>TV</b>	<b>163N</b> K <b>SK</b>		
Chimp	<i>Pan troglodytes</i>				<b>79N</b> SSN	<b>102N</b> V <b>TQ</b>					<b>134N</b> F <b>TV</b>	<b>163N</b> K <b>SK</b>		
Orangutan	<i>Pongo abelii</i>				<b>79N</b> SSN	<b>102N</b> V <b>TQ</b>					<b>134N</b> F <b>TV</b>	<b>163N</b> K <b>SK</b>		
Gibbon	<i>Nomascus leucogenys</i>				<b>79N</b> SSN	<b>102N</b> V <b>TQ</b>					<b>134N</b> F <b>TV</b>	<b>163N</b> K <b>SK</b>		
Rhesus	<i>Macaca mulatta</i>				<b>79N</b> SSN	<b>102N</b> I <b>TQ</b>					<b>134N</b> F <b>TV</b>	<b>163N</b> K <b>SK</b>		
Marmoset	<i>Callithrix jacchus</i>				<b>79N</b> SSN	<b>102N</b> V <b>TQ</b>					<b>134N</b> F <b>TV</b>			
Mouse	<i>Mus musculus</i>				<b>79N</b> SS <b>K</b>	<b>102N</b> I <b>TQ</b>					<b>134N</b> F <b>TV</b>			
Rat	<i>Rattus norvegicus</i>				<b>79N</b> SS <b>K</b>	<b>102N</b> I <b>TQ</b>					<b>134N</b> F <b>TV</b>			
Guinea Pig	<i>Cavia porcellus</i>				<b>79N</b> SSN	<b>102N</b> V <b>TQ</b>					<b>132N</b> D <b>TF</b>			<b>172N</b> R <b>TL</b>
Cow	<i>Bos taurus</i>				<b>79N</b> SS <b>K</b>	<b>102N</b> I <b>TQ</b>								<b>172N</b> R <b>TL</b>
Horse	<i>Equus caballus</i>				<b>79N</b> SS <b>K</b>	<b>102N</b> I <b>H</b>	<b>109N</b> H <b>TV</b>				<b>134N</b> D <b>TF</b>			
Dog	<i>Canis familiaris</i>				<b>79N</b> SS <b>K</b>	<b>102N</b> I <b>H</b>								<b>172N</b> R <b>TV</b>
Pig	<i>Sus scrofa</i>				<b>79N</b> SS <b>V</b>	<b>102N</b> I <b>TQ</b>					<b>132N</b> D <b>TF</b>			
Rabbit	<i>Oryctolagus cuniculus</i>				<b>79N</b> SSN	<b>102N</b> V <b>TQ</b>					<b>132N</b> D <b>TF</b>			
Elephant	<i>Loxodonta africana</i>				<b>79N</b> SSN	<b>102N</b> I <b>TQ</b>					<b>132N</b> D <b>TF</b>			
Panda	<i>Ailuropoda melanoleuca</i>				<b>79N</b> SS <b>A</b>	<b>102N</b> I <b>H</b>					<b>132N</b> D <b>TL</b>			
Opossum	<i>Monodelphis domestica</i>				<b>79N</b> SS <b>T</b> K	<b>102N</b> L <b>TQ</b>					<b>131N</b> D <b>SF</b>			
Platypus	<i>Ornithorhynchus anatinus</i>				<b>79N</b> SS <b>K</b>	<b>102N</b> I <b>T</b> K								
Chicken	<i>Gallus gallus</i>	<b>46N</b> G <b>TI</b>	72N <b>P</b> S <b>D</b>			<b>102N</b> I <b>T</b> E	<b>108N</b> G <b>TI</b>				<b>131N</b> D <b>TI</b>			<b>171N</b> R <b>TV</b>
Zebra finch	<i>Taeniopygia guttata</i>	<b>46N</b> G <b>G</b> T	72N <b>P</b> S <b>E</b>			<b>102N</b> V <b>T</b> E	<b>108N</b> G <b>TI</b>				<b>131N</b> D <b>TL</b>			<b>171N</b> R <b>TV</b>
Lizard	<i>Anolis carolensis</i>	<b>46N</b> G <b>TI</b>			<b>79N</b> G <b>S</b> Q	<b>102N</b> I <b>H</b>					<b>131N</b> D <b>TF</b>			
Frog	<i>Xenopus tropicalis</i>					<b>101N</b> I <b>T</b> Q	<b>107N</b> N <b>T</b> V						<b>162N</b> S <b>SL</b>	
Zebrafish	<i>Danio rerio</i>	<b>47N</b> G <b>T</b> L				<b>103N</b> I <b>T</b> F	<b>109N</b> N <b>T</b> V							<b>168N</b> R <b>TV</b>
Tetraodon	<i>Tetraodon nigroviridis</i>				<b>77N</b> G <b>T</b> T	<b>100N</b> V <b>T</b> Y	<b>105N</b> D <b>S</b> T						<b>162N</b> S <b>SL</b>	
Sea squirt	<i>Ciona intestinalis</i>			<b>74N</b> V <b>T</b> N			<b>120N</b> K <b>T</b> Y		<b>143N</b> G <b>S</b> E					
Lancelet	<i>Branchiostoma floridae</i>					<b>100N</b> I <b>T</b> F	<b>106N</b> G <b>T</b> V	<b>122N</b> M <b>S</b> F		<b>129N</b> D <b>TF</b>				
Fruit fly	<i>Drosophila melanogaster</i>			<b>80N</b> V <b>T</b> N	<b>90N</b> G <b>S</b> K		<b>118N</b> G <b>T</b> L							



Table 2. Cont.

Vertebrate	Species	Site 14	Site 15	Site 16	Site 17	Site 18	Site 19	Site 20	Site 21	Site 22	Site 23	Site 24	Site 25	No of Sites
<b>CD36</b>														
<b>Human</b>	<i>Homo sapiens</i>		205NNTA		220NISK		235NLSY		247NGTD		321NCTS		417NETG	10
<b>Chimp</b>	<i>Pan troglodytes</i>		205NNTA		220NISK		235NLSY		247NGTD		321NCTS		417NETG	10
<b>Orangutan</b>	<i>Pongo abelii</i>		205NNTA				235NLSY		247NGTD		321NCTS		417NETG	9
<b>Gibbon</b>	<i>Nomascus leucogenys</i>		205NNTA				235NLSY		247NGTD		321NCTS		417NETG	9
<b>Rhesus</b>	<i>Macaca mulatta</i>		205NNTA				235NLSY		247NGTD		321NCTS		417NETG	9
<b>Marmoset</b>	<i>Callithrix jacchus</i>		205NNTA				235NLSY		247NGTD		321NCTS		417NETG	9
<b>Mouse</b>	<i>Mus musculus</i>		205NNTV		220NISK		235NLSY		247NGTD		321NCTS		417NETG	9
<b>Rat</b>	<i>Rattus norvegicus</i>		205NNTV		220NISK		235NLSY		247NGTD		321NCTS		417NETG	9
<b>Guinea Pig</b>	<i>Cavia porcellus</i>		205NNTA		220NISK		235NLSY		247NGTD		321NCTS		417NETG	10
<b>Cow</b>	<i>Bos taurus</i>		205NNTA				235NLSY		247NGTD		321NCTS		417NETG	8
<b>Horse</b>	<i>Equus caballus</i>		205NNTV		220NISK		235NLSY		247NGTD		321NCTS		417NETG	10
<b>Dog</b>	<i>Canis familiaris</i>		205NNTV		220NVSQ		235NLSY		247NGTD		321NCTS		417NETG	9
<b>Pig</b>	<i>Sus scrofa</i>		205NNTS	206NTSD			235NLSY		247NGTD		321NCTS		417NETG	9
<b>Rabbit</b>	<i>Oryctolagus cuniculus</i>		205NNTV		220NISK		235NLSY		247NGTD		321NCTS		417NETG	9
<b>Elephant</b>	<i>Loxodonta africana</i>		205NNTV				235NLSY		247NGTD		321NCTS		417NETG	8
<b>Panda</b>	<i>Ailuropoda melanoleuca</i>		208NNTA				238NLSY		250NGTD		324NCTS		420NETG	8
<b>Opossum</b>	<i>Monodelphis domestica</i>		204NNTV				234NLSF		246NGTD		320NCTS		416NETG	8
<b>Platypus</b>	<i>Ornithorhynchus anatinus</i>		204NNTA				234NLSY		246NGTD		320NCTS		416NETG	7
<b>Chicken</b>	<i>Gallus gallus</i>		204NGTS				234NLSY		246NGTD		320NCTL		416NETA	10
<b>Zebra finch</b>	<i>Taeniopygia guttata</i>		205NGTS						247NGTD		321NCTI		417NESA	9
<b>Lizard</b>	<i>Anolis carolensis</i>		205NETL				232NKSM		247NTGD		321NCTG		417NETA	9
<b>Frog</b>	<i>Xenopus tropicalis</i>		202NGTA						245NGTD		319NCTA		415NETA	7
<b>Zebrafish</b>	<i>Danio rerio</i>		194NGTV					229NDSY	237NGSD		311NCTL		406NETA	9
<b>Tetraodon</b>	<i>Tetraodon nigroviridis</i>		200NGTA					228NRTV	243NGTD		317NCTL		416NETA	9
<b>Sea squirt</b>	<i>Ciona intestinalis</i>		232NQSR			260NMSE			276NGTD	346NHTV				7
<b>Lancelet</b>	<i>Branchiostoma floridae</i>	182NDSL	211NGTD						255NGTD			333NISI	420NEST	9
<b>Fruit fly</b>	<i>Drosophila melanogaster</i>		223NGTS									347NVSL		5

#### 2.4. Conserved Vertebrate CD36 Cysteine Residues

Ten cysteine residues of the vertebrate CD36 sequences were conserved, including two within each of the *N*- (Cys3 and Cys7) and *C*-terminal (Cys464 and Cys466) cytoplasmic sequences, and six within the vertebrate exoplasmic sequences (Cys243; Cys272; Cys311; Cys313; Cys322; and Cys333) (Figure 1). The CD36 *N*- and *C*-terminal conserved cytoplasmic cysteine residues have been shown to be palmitoylated [47], which may contribute to protein-protein interactions, protein trafficking and membrane localization [54]. Comparative studies of vertebrate SCARB1 sequences have shown that *N*- and *C*-terminal cytoplasmic sequences lacked any conserved cysteine residues in this region [55]. The six conserved exoplasmic vertebrate CD36 cysteine residues participate in disulfide bridge formation for bovine CD36 (Cys243-Cys311; Cys272-Cys333; and Cys313-Cys322), resulting in a 1-3, 2-6 and 4-5 arrangement of the disulfide bridges [49]. In contrast, vertebrate SCARB1 exoplasmic sequences contain only four conserved cysteine residues forming disulfide bridges (Cys281; Cys321; Cys323; and Cys334); a fifth cysteine (Cys251) was not conserved among vertebrate SCARB1 sequences [55]; and a conserved sixth cysteine (not observed in the CD36 sequence) (human SCARB1 Cys384) which functions in lipid transfer activity [56,57].

#### 2.5. Predicted Secondary Structures for Vertebrate CD36

Predicted secondary structures for vertebrate CD36 sequences were examined (Figure 1), particularly for the exoplasmic sequences.  $\alpha$ -Helix and  $\beta$ -sheet structures were similar in each case, with a  $\alpha$ -helix extending beyond the *N*-terminal and *C*-terminal trans-membrane regions, forming  $\alpha$ 1 and  $\alpha$ 7, respectively. A consistent sequence of predicted secondary structure was observed for each of the vertebrate CD36 sequences: *N*-terminal cytoplasmic sequence--*N*-terminal transmembrane sequence-- $\alpha$ 1-- $\beta$ 1-- $\alpha$ 2-- $\beta$ 2-- $\beta$ 3-- $\beta$ 4-- $\beta$ 5-- $\beta$ 6-- $\alpha$ 3-- $\beta$ 7-- $\alpha$ 4-- $\beta$ 8-- $\beta$ 10-- $\beta$ 11-- $\beta$ 12-- $\beta$ 13-- $\beta$ 14-- $\beta$ 15-- $\alpha$ 5--*C*-terminal trans-membrane sequence--*C*-terminal cytoplasmic sequence. Further description of the secondary and tertiary structures for CD36 must await the determination of the three dimensional structure for this protein, particularly for the exoplasmic region which directly binds oxidized LDL lipids and a wide range of other lipid-like structures, including long chain fatty acids [1–10].

#### 2.6. Conserved Proline, Glycine and Charged Amino Acid Residues within the CD36 Exoplasmic Domain

Supplementary Figure 1 shows the alignment of 7 vertebrate CD36 amino acid sequences for the exoplasmic domain with colors depicting the properties of individual amino acids and conservation observed for some of these protein sequences. In addition to the key vertebrate CD36 amino acids detailed previously, others were also conserved, including 17 proline residues. A human CD36 genetic deficiency of one of these conserved prolines (Pro90→Ser) confirmed the significance of this residue, which lacked platelet CD36 [56]. Human CD36 deficiency has been shown to cause systemic metabolic changes in glucose and long chain fatty acid metabolism [59]. Prolines play a major role in protein folding and protein-protein interactions, involving the cyclic pyrrolidine amino acid side chain, which may introduce turns (or kinks) in the polypeptide chain as well as having destabilizing effects on  $\alpha$ -helix and  $\beta$ -strand conformations [60]. In addition, the presence of sequential prolines within a protein sequence may confer further restriction in folding conformation and create a distinctive

structure, such as that reported for the mammalian  $\text{Na}^+/\text{H}^+$  exchanger, which plays a major role in cation transport [61]. Sequential prolines (Pro258-Pro259) were conserved for 6 of 7 vertebrate CD36 sequences examined and these may confer a distinctive conformation in this region supporting the lipid receptor functions for this protein. Moreover, regions of water exposed proteins with high levels of proline residues are often sites for protein-protein interactions [62] and these residues may significantly contribute to the binding of lipoproteins by the exoplasmic region of CD36. Similar results have been recently reported for the vertebrate SCARB1 exoplasmic region, however in this case, 30 conserved proline residues were observed [55].

Supplementary Figure 1 also shows conservation of 14 glycine residues for vertebrate CD36 exoplasmic domains, which due to their small size, may be essential for static turns, bends or close packing in the domain, or required for conformational dynamics during long chain fatty acid receptor on-off switching, as in the case of the aspartate receptor protein [63]. Both proline and glycine residues are frequently found in turn and loop structures of proteins, and usually influence short loop formation within proteins containing between 2 and 10 amino acids [61]. Evidence for these short loop structures within vertebrate CD36 exoplasmic sequences was evident from the predicted secondary structures for vertebrate CD36 (Figure 1), with proline and/or glycine residues found at the start of the following structures:  $\alpha 1$  (Pro28; Gly30),  $\beta 1$  (Gly58),  $\alpha 2$  (Pro73),  $\beta 3$  (Gly89-Pro90),  $\beta 8$  (Gly210),  $\beta 12$  (Gly287) and  $\alpha 5$  (Gly420; Gly423). Moreover, CD36 sequential proline residues (Pro255-Pro256) were located in a region with no predicted secondary structure (between  $\beta 9$  and  $\beta 10$ ) but with disulfide bonds, which suggests that this is a region of conformational significance for CD36.

In addition to the prolines and glycine residues for the vertebrate exoplasmic CD36 sequences, there are several conserved charged amino acid residue positions, including positively charged Lys40/Lys41 located within the first predicted exoplasmic helix ( $\alpha 1$ ); Arg/Lys89, Arg95; Arg97 and Lys101 within or near the predicted strand- $\beta 3$ /strand  $\beta 4$  THP-binding domain region; Lys233/Lys235/Arg236 near the PE-binding domain; Lys263 located near the  $\beta 10$  strand; Arg276 within the  $\beta 11$  strand and adjacent to a disulphide bond; Lys288 which lies between predicted  $\beta 11$  and  $\beta 12$  strands; Lys337 and Arg/Lys340 near a disulphide bond; Lys388/Arg389 near the predicted  $\beta 15$  strand; and Lys401/Lys409 within the last exoplasmic helix ( $\alpha 5$ ). Two domains of the exoplasmic CD36 sequence have been potentially implicated in the binding and endocytosis of apoptotic neutrophils: residues 155-183; and 93-120 (see [7]) The latter domain is called CLESH (for CD36 LIMP-II Emp [erythrocyte membrane protein] sequence homology) which is predominantly conserved, particularly near Thr92, which is phosphorylated by protein kinase C alpha and contributes to the suppression of thrombospondin-1 binding *in vitro* [48]. One or more of these positively charged CD36 exoplasmic regions may contribute to long chain fatty acid binding prior to the translocation of fatty acids inside the cell membrane. There are also several conserved acidic amino acid regions, particularly a sequence of three acidic amino acids (367Asp/68Asp/369Asp) near the  $\beta 13$  predicted strand. The conserved nature of these CD36 charged residues suggests that play key functional roles for this cell membrane protein, which may include serving as the long chain fatty acid CD36 receptor site.

### 2.7. Alignments of Human CD36, SCARB1 and SCARB2

The amino acid sequences for human CD36, SCARB1 and SCARB2 (see Table 1) are aligned in Figure 2. The sequences were 30-33% identical and showed similarities in several key features and residues, including cytoplasmic *N*-terminal and *C*-terminal residues; *N*-terminal and *C*-terminal trans-membrane helical regions; exoplasmic disulfide bond forming residues, previously identified for bovine CD36: Cys243-Cys311; Cys272-Cys333; and Cys313-Cys322 [47]; several predicted *N*-glycosylation sites for human CD36 (10 sites), SCARB1 (9 sites) and SCARB2 (9 sites), of which only two are shared between these sequences (*N*-glycosylation sites 15 and 21 (Table 2); and similar predicted secondary structures previously identified for SCARB1 [55] (Figure 1). The Cys384 residue, for which the free-SH group plays a major role in SCARB1-mediated lipid transport [57], was unique to SCARB1, being replaced by other residues for the corresponding CD36 and SCARB2 proteins (Phe383 and Ala379, respectively). *N*-terminal trans-membrane glycine residues, which play a role in the formation of SCARB1 oligomers [53], were also observed for the human CD36 sequence, with twin-glycines (Gly23-Gly24) conserved for the vertebrate CD36 sequences (Figure 1). In contrast, only one of these glycines (Gly10) was observed for the human SCARB2 sequence. These results suggest that human CD36, SCARB1 and SCARB2 proteins share several important properties, features and conserved residues, including being membrane-bound with cytoplasmic and transmembrane regions, having similar secondary structures, but being significantly different to serve distinct functions.

Alignments were also prepared for the predicted lancelet (*Branchiostoma floridae*) and sea squirt (*Ciona intestinalis*) CD36-like sequences and a major epithelial membrane protein (EMP) from fruit fly (*Drosophila melanogaster*) (FBpp0072309) with the human CD36, SCARB1 and SCARB2 sequences (Figure 2). The lancelet, sea squirt and fruit fly sequences examined shared many features with the CD36-like human sequences, including the *N*- and *C*-terminal cytoplasmic and transmembrane sequences; similarities in predicted secondary structures; positional identities for five conserved cysteine residues, indicating conservation of at least 2 disulfide bridges for these proteins; predicted *N*-glycosylation sites, including one which is shared across all 6 CD-like sequences (site 15 in Table 2); and trans-membrane glycine residues, which were observed in both the *N*- and *C*-terminal sequences.

### 2.8. Gene Locations and Exonic Structures for Vertebrate CD36 Genes

Table 1 summarizes the predicted locations for vertebrate *CD36* genes based upon BLAT interrogations of several vertebrate genomes using the reported human *CD36* sequence [45] and the predicted sequences for other vertebrate *CD36* genes and the UC Santa Cruz genome browser [64]. Vertebrate *CD36* genes were transcribed on either the positive strand (e.g., human, chimpanzee, gibbon, rhesus, rat and dog genomes) or the negative strand (e.g., mouse, cow, pig, opossum, chicken, frog and zebrafish genomes). Figure 1 summarizes the predicted exonic start sites for human, mouse, cow, opossum, chicken, frog and zebrafish *CD36* genes with each having 12 coding exons, in identical or similar positions to those reported for the human *CD36* gene [28].

**Figure 2.** Amino Acid Sequence Alignments for Human CD36, SCARB1, and SCARB2; and Lancelet, Sea Squirt and Fruit Fly CD36-like Sequences. See Table 1 for sources of CD36-like sequences; HuCD36-human CD36; HuSCA1-human SCARB1; HuSCA2-human SCARB2; LaCD36-lancelet CD36; SsCD36- sea squirt CD36; DmEMP1-fruit fly endothelial membrane protein; \* shows identical residues for subunits; : similar alternate residues; . dissimilar alternate residues; predicted cytoplasmic residues are shown in red; predicted trans-membrane residues are shown in blue; N-glycosylated and potential N-glycosylated Asn sites are shown in green; free-SH Cys involved in lipid transfer for human SCARB1 is shown in pink; predicted disulfide bond Cys residues are shown in cyan; predicted  $\alpha$ -helices for CD36-like sequences are in shaded yellow and numbered in sequence from the start of the predicted exoplasmic domain; predicted  $\beta$ -sheets are in shaded grey and also numbered in sequence; bold underlined font shows residues corresponding to known or predicted exon start sites; exon numbers are shown; **G** residues refer to conserved glycines in the N- and C-terminal oligomerisation domains of the trans-membrane sequence [49]; C-terminal SCARB1 **AKL** residues refer to PDZ-binding domain sequences [18,19].

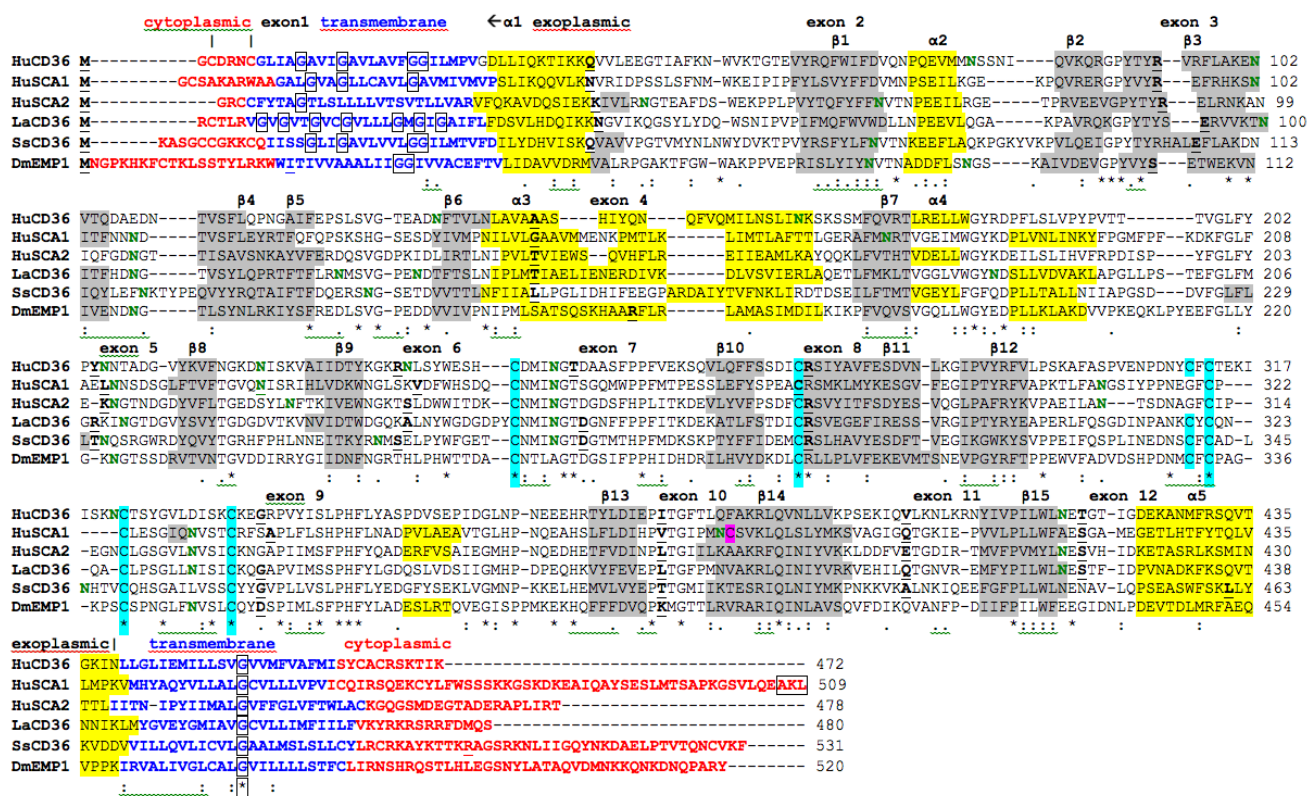
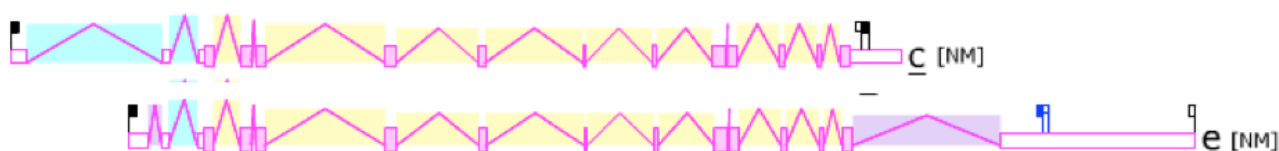


Figure 3 shows the predicted structures of mRNAs for two major human *CD36* transcripts and the major *Cd36* transcripts for mouse and rat *Cd36* genes [46,65,66]. The human *CD36* transcripts were ~2 kb in length with 14 (isoform c) or 15 (isoform e) introns present for these *CD36* mRNA transcripts and in each case, a 3'-untranslated region (UTR) was observed. The human *CD36* genome sequence contained a number of predicted transcription factor binding sites (TFBS), including the dual promoter structure of PPARA (peroxisome proliferator-activated receptor- $\alpha$ ) and PPARG (peroxisome proliferator-activated receptor- $\gamma$ ) sites [67,68]. Moreover, the mouse *Cd36* gene is regulated in a tissue

specific manner by PPARA in liver and by PPARG in adipose tissues [69]. Other TFBS sites predicted for the human CD36 5' promoter region included RSRFC4, a myocyte enhancer factor 2A found in muscle-specific and 'immediate early' genes [70]; CART1, a paired-class homeodomain transcription factor [71]; FOXJ2, a fork head transcriptional activator which is active during early development [72]; XBP1, a transcription factor which is critical for cell fate determination in response to endoplasmic reticulum stress [73]; and CDC5, a transcription activator and cell cycle regulator [74]. Hepatic upregulation of CD36 transcription in human patients has been recently shown to be significantly associated with insulin resistance, hyperinsulinaemia and increased steatosis in non-alcoholic steatohepatitis and chronic hepatitis C [43].

**Figure 3.** Gene Structures and Major Splicing Transcripts for the Human, Mouse and Rat *CD36* Genes. Derived from the AceView website <http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/> mature isoform variants are shown with capped 5'- and 3'- ends for the predicted mRNA sequences [62]; NM refers to the NCBI reference sequence; exons are in pink; the directions for transcription are shown as 5' → 3'; sizes of mRNA sequences are shown in kilobases (kb); predicted transcription factor binding sites (TFBS) for human *Cd36* are shown: CART1- a paired-class homeodomain transcription factor [71]; RSRFC4-myocyte enhancement factor 2A transcription factor [70]; XBP1-transcription factor [73]; FOXJ2-fork-head transcription factor[72]; CDC5-transcription activator and cell cycle regulator; [74]; PPARA-peroxisome proliferator-activated receptor alpha; and PPARG-peroxisome proliferator-activated receptor gamma [67,68].

Human *CD36* 5'→3' chromosome 7:79,836,828-80,146,532 size=309.7kb on plus strand  
6.6 times average gene expression level



**Predicted transcription factor binding sites**

CART1 RSRFC4 XBP1  
FOXJ2 CDC5 PPARA PPARG

Mouse *Cd36* 5'→3' chromosome 5:17,394,781 to 17,287,508 size=107.3kb on minus strand  
4.2 times average gene expression level



Rat *Cd36* 5'→3' chromosome 4:13,065,071 to 13,525,617 size=60.5kb on plus strand  
0.3 times average gene expression level



2.9. Comparative Human and Mouse CD36 Tissue Expression

Figure 4 presents ‘heat maps’ showing comparative gene expression for various human and mouse tissues obtained from GNF Expression Atlas Data using the U133A and GNF1H (human) and GNF1M (mouse) chips (<http://genome.ucsc.edu>; <http://biogps.gnf.org>) [75]. These data supported a broad and high level of tissue expression for human and mouse *CD36*, particularly for adipose tissue, heart, skeletal muscle and liver, which is consistent with previous reports for these genes [11,32,66]. Overall, human and mouse *CD36* tissue expressions levels were 4-6 times the average level of gene expression which supports the key role played by this enzyme in fatty acid metabolism, especially in liver, muscle and adipose tissue.

**Figure 4.** Comparative Tissue Expression for Human and Mouse *CD36* Genes. Expression ‘heat maps’ (GNF Expression Atlas 2 data) (<http://biogps.gnf.org>) were examined for comparative gene expression levels among human and mouse tissues for *CD36* genes showing high (red); intermediate (black); and low (green) expression levels [75]. Derived from human and mouse genome browsers (<http://genome.ucsc.edu>) [64].



The broad tissue and high level of gene expression reported for human and mouse *CD36* reflects key roles for this major cell membrane and muscle outer mitochondrial membrane glycoprotein in fatty acyl translocation and as a multiple ligand cell surface receptor of oxidized LDL lipoproteins (ox-LDL) and long chain fatty acids [7,11,33,66]. *CD36* has also been described as a lipid ‘sensor’ playing a lipid receptor role for cells and tissues of the body [8,40]. Moreover, *CD36* upregulation is associated with insulin resistance and hyperinsulinaemia, leading to liver pathology and increased steatosis [43]. In addition, cardiomyocyte *CD36* cell surface recruitment is induced by insulin, AMP-dependent protein kinase (AMPK) activity or contraction, and is regulated in its vesicular trafficking by the RabGAP-AS160 substrate and AS160-Rab8a GTPase activating protein (GAP) [76–78]. These features provide a link between cell membrane *CD36* and the reported insulin-stimulated

phosphorylation of AS160 involved with the translocation of the glucose transporter GLUT4 to the plasma membrane [79,80]. It is also relevant to report that plasma levels of soluble CD36 are increased in type 2 diabetic patients [81].

Significant levels of CD36 expression have also been described in brain tissues, where CD36 contributes to cerebrovascular oxidative stress and neurovascular dysfunction induced by amyloid-beta in Alzheimer's dementia [12,13], and in transporting long chain fatty acids across the blood-brain barrier [82].

### 2.10. Phylogeny of Vertebrate CD36-Like Sequences

A phylogenetic tree (Figure 5) was calculated by the progressive alignment of 21 vertebrate CD36 amino acid sequences with human, mouse, chicken and zebrafish SCARB1 and SCARB2 sequences. The phylogenetic tree was 'rooted' with the lancelet (*Branchiostoma floridae*) CD36 sequence (see Table 1). The phylogenetic tree showed clustering of the CD36 sequences into groups which were consistent with their evolutionary relatedness as well as groups for human, mouse, chicken and zebrafish SCARB1 and SCARB2 sequences, which were distinct from the lancelet CD36 sequence. These groups were significantly different from each other (with bootstrap values of ~100/100) with the clustering observed supporting a closer phylogenetic relationship between *CD36* and *SCARB2*, with the *SCARB1* gene being more distantly related. This is suggestive of a sequence of *CD36*-like gene duplication events: ancestral *CD36* gene duplication → *SCARB1* and *CD36* genes; followed by a further *CD36* duplication, generating the *SCARB2* and *CD36* genes found in all vertebrate species examined (Figure 5). It is apparent from this study of vertebrate *CD36*-like genes and proteins that this is an ancient protein for which a proposed common ancestor for the *CD36*, *SCARB1* and *SCARB2* genes may have predated the appearance of fish > 500 million years ago [83]. In parallel with the evolution of *CD36* and other *CD36*-like proteins (*SCARB1* and *SCARB2*), thrombospondins (TSPs) are also undergoing evolutionary changes in their structures and functions [84], with gene duplication events proposed at the origin of deuterostomes.

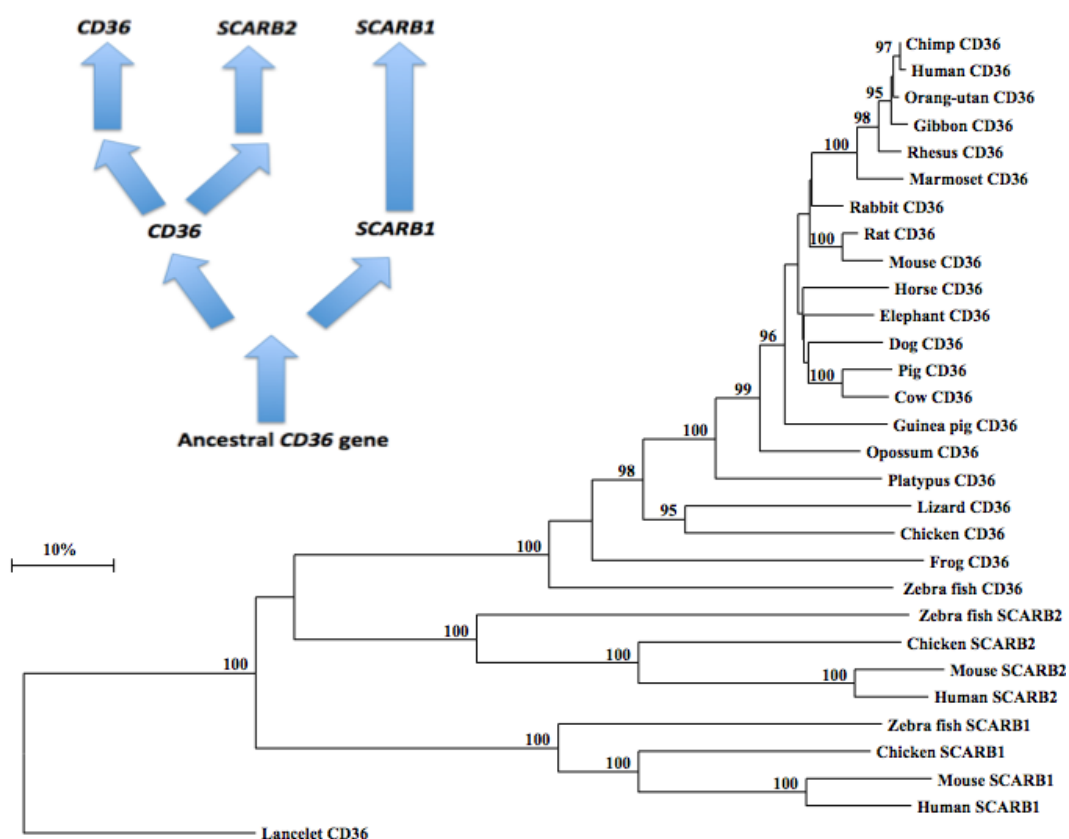
## 3. Methods

### 3.1. Vertebrate CD36 Gene and Protein Identification

BLAST (Basic Local Alignment Search Tool) studies were undertaken using web tools from the National Center for Biotechnology Information (NCBI) (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) [85]. Protein BLAST analyses used vertebrate CD36 amino acid sequences previously described [8,45] (Table 1). Non-redundant protein sequence databases for several vertebrate genomes were examined using the blastp algorithm from sources previously described [55]. This procedure produced multiple BLAST 'hits' for each of the protein databases which were individually examined and retained in FASTA format, and a record kept of the sequences for predicted mRNAs and encoded CD36-like proteins. Predicted CD36-like protein sequences were obtained in each case and subjected to analyses of predicted protein and gene structures.



**Figure 5.** Phylogenetic Tree of Vertebrate CD36 Amino Acid Sequences with Human, Mouse, Chicken and Zebrafish SCARB1 and SCARB2 Sequences. The tree is labeled with the *CD36*-like gene name and the name of the animal and is ‘rooted’ with the lancelet CD36 sequence. Note the 3 major clusters corresponding to the *CD36*, *SCARB1* and *SCARB2* gene families. A genetic distance scale is shown. The number of times a clade (sequences common to a node or branch) occurred in the bootstrap replicates are shown. Only highly significant replicate values of 95 or more are shown with 100 bootstrap replicates performed in each case. A proposed sequence of *CD36* gene duplication events is shown.



BLAT (Blast-like Alignment Tool) analyses were subsequently undertaken for each of the predicted CD36 amino acid sequences using the UC Santa Cruz Genome Browser [64] with the default settings to obtain the predicted locations for each of the vertebrate *CD36* genes, including predicted exon boundary locations and gene sizes. BLAT analyses were similarly undertaken for vertebrate *SCARB1* and *SCARB2* genes using previously reported sequences in each case (see Table 1). Structures for human and mouse isoforms (splicing variants) for human *CD36*, mouse *Cd36* and rat *Cd36* were obtained using the AceView website to examine predicted gene and protein structures [66].

### 3.2. Predicted Structures and Properties of Vertebrate CD36

Predicted secondary structures for vertebrate CD36 proteins, human SCARB1 and SCARB2, lancelet (*Branchiostoma floridae*) CD36, sea squirt (*Ciona intestinalis*) CD36 and a fruit fly (*Drosophila melanogaster*) epithelial membrane protein (FBpp0072309) were obtained using the

PSIPRED v2.5 web site tools provided by Brunel University [86]. Molecular weights, *N*-glycosylation sites [49] and predicted trans-membrane, cytosolic and exocellular sequences for vertebrate SCARB1 proteins were obtained using Expasy web tools ([http://au.expasy.org/tools/pi\\_tool.html](http://au.expasy.org/tools/pi_tool.html)).

### 3.3. Comparative Human and Mouse CD36 Gene Expression

The genome browser (<http://genome.ucsc.edu>) [62] was used to examine GNF Expression Atlas 2 data using various expression chips for human and mouse *CD36* genes (<http://biogps.gnf.org>) [74]. Gene array expression ‘heat maps’ were examined for comparative gene expression levels among human and mouse tissues showing high (red); intermediate (black); and low (green) expression levels.

### 3.4. Phylogeny Studies and Sequence Divergence

Alignments of vertebrate CD36, SCARB1 and SCARB2 sequences were assembled using BioEdit v.5.0.1 and the default settings [87]. Alignment ambiguous regions, including the amino and carboxyl termini, were excluded prior to phylogenetic analysis yielding alignments of 431 residues for comparisons of vertebrate CD36 sequences with human, mouse, chicken and zebra-fish SCARB1 and SCARB2 sequences with the lancelet (*Branchiostoma floridae*) CD36 sequence (Table 1). Evolutionary distances and phylogenetic trees were calculated as previously described [85]. Tree topology was reexamined by the boot-strap method (100 bootstraps were applied) of resampling and only values that were highly significant ( $\geq 95$ ) are shown [88].

## 4. Conclusions

The results of this study indicate that vertebrate *CD36* genes and encoded proteins represent a distinct gene and protein family of *CD36*-like proteins which share key conserved sequences that have been reported for other *CD36*-like proteins (SCARB1 and SCARB2) previously studied [16–24]. *CD36* has a unique property among these proteins in serving a major role in fatty acyl translocation and as a multiple ligand cell surface receptor of oxidized LDL lipoproteins (ox-LDL), long chain fatty acids, aged neutrophils and *Plasmodium falciparum*-parasitized erythrocytes [3–10]. *CD36* is encoded by a single gene among the vertebrate genomes studied and is highly expressed in human and mouse tissues, particularly in adipose tissue, heart, skeletal muscle and liver, and usually contain 12 coding exons. Predicted secondary structures for vertebrate *CD36* proteins showed strong similarities with other *CD36*-like proteins, SCARB1 and SCARB2. Three major structural domains were observed for vertebrate *CD36* sequences, including *N*- and *C*-terminal cytoplasmic domains; *N*- and *C*-terminal trans-membrane domains; and an exoplasmic domain, which serves as the ‘receptor’ for long chain fatty acids and thrombospondins [5–8,14,32]. The latter domain contained three disulfide bridges [49]; several *N*-glycosylation sites for glycan binding (7–10 sites), which are essential for membrane recruitment [52]; 17 conserved proline and 14 glycine residues, which may contribute to short loop structures for the *CD36* exoplasmic structure; and several conserved basic amino acid sites, which may promote long chain fatty acid binding. Phylogenetic studies using 21 vertebrate *CD36* sequences with human, mouse, chicken and zebrafish SCARB1 and SCARB2 sequences indicated that the *CD36* gene appeared early in evolution, prior to the appearance of bony fish more than 500 million years ago, and

has undergone at least two gene duplication events: ancestral *CD36* → vertebrate *SCARB1* and *CD36*; with the latter gene undergoing a further gene duplication generating vertebrate *CD36* and *SCARB2* genes.

### Acknowledgements

The expert assistance of Bharet Patel of Griffith University with the phylogeny studies is acknowledged.

### References

1. Oquendo, P.; Hundt, E.; Lawler, J.; Seed, B. CD36 directly mediates cytoadherence of *Plasmodium falciparum* parasitized erythrocytes. *Cell* **1989**, *58*, 95–101.
2. Asch, A.S.; Silbiger, S.; Heimer, E.; Nachman, R.L. Thrombospondin sequence motif (CSVTCG) is responsible for CD36 binding. *Biochem. Biophys. Res. Commun.* **1992**, *182*, 1208–1217.
3. Navazo, M.D.; Daviet, L.; Savill, J.; Ren, Y.; Leung, L.L.; McGregor, J.L. Identification of a domain (155–183) on CD36 implicated in the phagocytosis of apoptotic neutrophils. *J. Biol. Chem.* **1996**, *271*, 15381–15385.
4. Serghides, L.; Crandall, I.; Hull, E.; Kain, K.C. The *Plasmodium falciparum*-CD36 interaction is modified by a single amino acid substitution in CD36. *Blood* **1998**, *92*, 1814–1819.
5. Simantov, R.; Silverstein, R.L. CD36: A critical anti-angiogenic receptor. *Front. Biosci.* **2003**, *8*, s874–s882.
6. Adachi, H.; Tsujimoto, M. Endothelial scavenger receptors. *Prog. Lipid Res.* **2006**, *45*, 379–404.
7. Collot-Teixeira, S.; Martin, J.; McDermott-Roe, C.; Poston, R.; McGregor, J.L. CD36 and macrophages in atherosclerosis. *Cardiovasc. Res.* **2007**, *75*, 468–477.
8. Martin, C.A.; Longman, E.; Wooding, C.; Hoosdally, S.J.; Ali, S.; Aitman, T.J.; Gutmann, D.A.; Freemont, P.S.; Byrne, B.; Linton, K.J. Cd36, a class B scavenger receptor, functions as a monomer to bind acetylated and oxidized low-density lipoproteins. *Protein Sci.* **2007**, *16*, 2531–2541.
9. Gautam, S.; Banerjee, M. The macrophage Ox-LDL receptor, CD36 and its association with type II diabetes mellitus. *Mol. Genet. Metab.* **2011**, *102*, 389–398.
10. Ren, Y. Peroxisome proliferator-activator receptor  $\gamma$ : A link between macrophage CD36 and inflammation in malaria infection. *PPAR Res.* **2012**, *2012*, 640769.
11. Smith, B.K.; Jain, S.S.; Rimbaud, S.; Dam, A.; Quadrilatero, J.; Ventura-Clapier, R.; Bonen, A.; Holloway, G.P. FAT/CD36 is located on the outer mitochondrial membrane, upstream of long-chain acyl-CoA synthetase, and regulates palmitate oxidation. *Biochem. J.* **2011**, *437*, 125–134.
12. Baranova, I.N.; Bocharov, A.V.; Vishnyakova, T.G.; Kurlander, R.; Chen, Z.; Fu, D.; Arias, I.M.; Csako, G.; Patterson, A.P.; Eggerman, T.L. CD36 is a novel serum amyloid A (SAA) receptor mediating SAA binding and SAA-induced signaling in human and rodent cells. *J. Biol. Chem.* **2010**, *285*, 8492–8506.
13. Park, L.; Wang, G.; Zhou, P.; Zhou, J.; Pitstick, R.; Previti, M.L.; Younkin, L.; Younkin, S.G.; Van Nostrand, W.E.; Cho, S.; *et al.* Scavenger receptor CD36 is essential for the cerebrovascular oxidative stress and neurovascular dysfunction induced by amyloid-beta. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 5063–5068.
14. Martin, C.; Chevrot, M.; Poirier, H.; Passilly-Degrace, P.; Niot, I.; Besnard, P. CD36 as a lipid sensor. *Physiol. Behav.* **2011**, *105*, 36–42.

15. Chen, K.; Li, W.; Silverstein, R.L. A specific CD36-dependent signaling pathway is required for platelet activation by oxidized low-density lipoprotein. *Circ. Res.* **2008**, *102*, 1512–1519.
16. Acton, S.; Osgood, D.; Donoghue, M.; Corella, D.; Pocovi, M.; Cenarro, A.; Mozas, P.; Keilty, J.; Squazzo, S.; Woolf, E.A.; *et al.* Association of polymorphisms at the SR-BI gene locus with plasma lipid levels and body mass index in a white population. *Arterioscler. Thromb. Vasc. Biol.* **1999**, *19*, 1734–1743.
17. Bultel-Brienne, S.; Lestavel, S.; Pilon, A.; Laffont, I.; Tailleux, A.; Fruchart, J.C.; Siest, G.; Clavey, V. Lipid free apolipoprotein E binds to the class B Type I scavenger receptor I (SR-BI) and enhances cholesteryl ester uptake from lipoproteins. *J. Biol. Chem.* **2002**, *277*, 36092–36099.
18. Marsche, G.; Zimmermann, R.; Horiuchi, S.; Tandon, N.N.; Sattler, W.; Malle, E. Class B scavenger receptors CD36 and SR-BI are receptors for hypochlorite-modified low density lipoprotein. *J. Biol. Chem.* **2003**, *278*, 47562–47570.
19. Connelly, M.A.; Williams, D.L. Scavenger receptor BI: A scavenger receptor with a mission to transport high density lipoprotein lipids. *Curr. Opin. Lipid.* **2004**, *5*, 287–295.
20. Kent, A.P.; Stylianou, I.M. Scavenger receptor class B member 1 protein: Hepatic regulation and its effects on lipids, reverse cholesterol transport, and atherosclerosis. *Hep. Med.: Evid. Res.* **2011**, *3*, 29–44.
21. Fujita, H.; Takata, Y.; Kono, A.; Tanaka, Y.; Takahashi, T.; Himeno, M.; Kato, K. Isolation and sequencing of a cDNA clone encoding the 85 kDa human lysosomal sialoglycoprotein (hLGP85) in human metastatic pancreas islet tumor cells. *Biochem. Biophys. Res. Commun.* **1992**, *184*, 604–611.
22. Ogata, S.; Fukuda, M. Lysosomal targeting of Limp II membrane glycoprotein requires a novel Leu-Ile motif at a particular position in its cytoplasmic tail. *J. Biol. Chem.* **1994**, *269*, 5210–5217.
23. Tabuchi, N.; Akasaki, K.; Sasaki, T.; Kanda, N.; Tsuji, H. Identification and characterization of a major lysosomal membrane glycoprotein, LGP85/LIMP II in mouse liver. *J. Biochem.* **1997**, *122*, 756–763.
24. Kuronita, T.; Eskelinen, E.L.; Fujita, H.; Saftig, P.; Himeno, M.; Tanaka, Y. A role for the lysosomal membrane protein LGP85 in the biogenesis and maintenance of endosomal and lysosomal morphology. *J. Cell Sci.* **2002**, *115*, 4117–4131.
25. Lin, Y.W.; Lin, H.Y.; Tsou, Y.L.; Chitra, E.; Hsiao, K.N.; Shao, H.Y.; Liu, C.C.; Sia, C.; Chong, P.; Chow, Y.H. Human SCARB2-Mediated Entry and Endocytosis of EV71. *PLoS One* **2012**, *7*, e30507.
26. Taylor, K.T.; Tang, Y.; Sobieski, D.A.; Lipsky, R.H. Characterization of two alternatively spliced 5'-untranslated exons of the human CD36 gene in different cell types. *Gene* **1993**, *33*, 205–212.
27. Fernández-Ruiz, E.; Armesilla, A.L.; Sánchez-Madrid, F.; Vega, M.A. Gene encoding the collagen type I and thrombospondin receptor CD36 is located on chromosome 7q11.2. *Genomics* **1993**, *17*, 759–761.
28. Armesilla, A.L.; Vega, M.A. Structural organization of the gene for human CD36 glycoprotein. *J. Biol. Chem.* **1994**, *269*, 18985–18991.
29. Rać, M.E.; Safranow, K.; Poncyljusz, W. Molecular basis of human CD36 gene mutations. *Mol. Med.* **2007**, *13*, 288–296.

30. Simantov, R.; Febbraio, M.; Silverstein, R.L. The antiangiogenic effect of thrombospondin-2 is mediated by CD36 and modulated by histidine-rich glycoprotein. *Matrix Biol.* **2005**, *24*, 27–34.
31. Seimon, T.A.; Nadolski, M.J.; Liao, X.; Magallon, J.; Nguyen, M.; Feric, N.T.; Koschinsky, M.L.; Harkewicz, R.; Witztum, J.L.; Tsimikas, S.; *et al.* Atherogenic lipids and lipoproteins trigger CD36-TLR2-dependent apoptosis in macrophages undergoing endoplasmic reticulum stress. *Cell Metab.* **2010**, *12*, 467–482.
32. Nergiz-Unal, R.; Rademakers, T.; Cosemans, J.M.; Heemskerk, J.W. CD36 as a multiple-ligand signaling receptor in atherothrombosis. *Cardiovasc. Hematol. Agents Med. Chem.* **2011**, *9*, 42–55.
33. Ruiz-Valasco, N.; Dominguez, A.; Vega, M.A. Statins upregulated CD36 expression in human monocytes, an effect strengthened when combined with PPAR-gamma ligands. Putative contribution of Rho GTPases in statin-induced expression. *Biochem. Pharmacol.* **2004**, *67*, 303–313.
34. Liu, H.; Liu, Q.; Lei, H.; Li, X.; Chen, X. Inflammatory stress promotes lipid accumulation in the aorta and liver of SR-A/CD36 double knock-out mice. *Mol. Med. Rep.* **2010**, *3*, 1053–1061.
35. Steinbusch, L.K.; Luiken, J.J.; Vlasblom, R.; Chabowski, A.; Hoebers, N.T.; Coumans, W.A.; Vroegrijk, I.O.; Voshol, P.J.; Ouwens, D.M.; Glatz, J.F.; *et al.* Absence of fatty acid transporter CD36 protects against Western-type diet-related cardiac dysfunction following pressure overload in mice. *Am. J. Physiol. Endocrinol. Metab.* **2011**, *301*, E618–E627.
36. Kennedy, D.J.; Kashyap, S.R. Pathogenic role of scavenger receptor CD36 in the metabolic syndrome and diabetes. *Metab. Syndr. Relat. Disord.* **2011**, *9*, 239–245.
37. Lally, J.S.; Jain, S.S.; Han, X.X.; Snook, L.A.; Glatz, J.F.; Luiken, J.J.; McFarlan, J.; Holloway, G.P.; Bonen, A. Caffeine-stimulated fatty acid oxidation is blunted in CD36 null mice. *Acta Physiol.* **2011**, doi:10.1111/j.1748-1716.2011.02396.x.
38. Ma, X.; Bacci, S.; Mlynarski, W.; Gottardo, L.; Soccio, T.; Menzaghi, C.; Iori, E.; Lager, R.A.; Shroff, A.R.; Gervino, E.V.; *et al.* A common haplotype at the CD36 locus is associated with high free fatty acid levels and increased cardiovascular risk in Caucasians. *Hum. Mol. Genet.* **2004**, *13*, 2197–2205.
39. Yuasa-Kawase, M.; Masuda, D.; Yamashita, T.; Kawase, R.; Nakaoka, H.; Inagaki, M.; Nakatani, K.; Tsubakio-Yamamoto, K.; Ohama, T.; Matsuyama, A.; *et al.* Patients with CD36 Deficiency Are Associated with Enhanced Atherosclerotic Cardiovascular Diseases. *J. Atheroscler. Thromb.* **2011**, *19*, 263–275.
40. Keller, K.L.; Liang, L.C.; Sakimura, J.; May, D.; van Belle, C.; Breen, C.; Driggin, E.; Tepper, B.J.; Lanzano, P.C.; Deng, L.; *et al.* Common Variants in the CD36 Gene Are Associated With Oral Fat Perception, Fat Preferences, and Obesity in African Americans. *Obesity* **2012**, doi:10.1038/oby.2011.374.
41. Pain, A.; Urban, B.C.; Kai, O.; Casals-Pascual, C.; Shafi, J.; Marsh, K.; Roberts, D.J. A non-sense mutation in Cd36 gene is associated with protection from severe malaria. *Lancet* **2001**, *357*, 1502–1503.
42. Omi, K.; Ohashi, J.; Patarapotikul, J.; Hananantachai, H.; Naka, I.; Looareesuwan, S.; Tokunaga, K. CD36 polymorphism is associated with protection from cerebral malaria. *Am. J. Hum. Genet.* **2003**, *72*, 364–374.

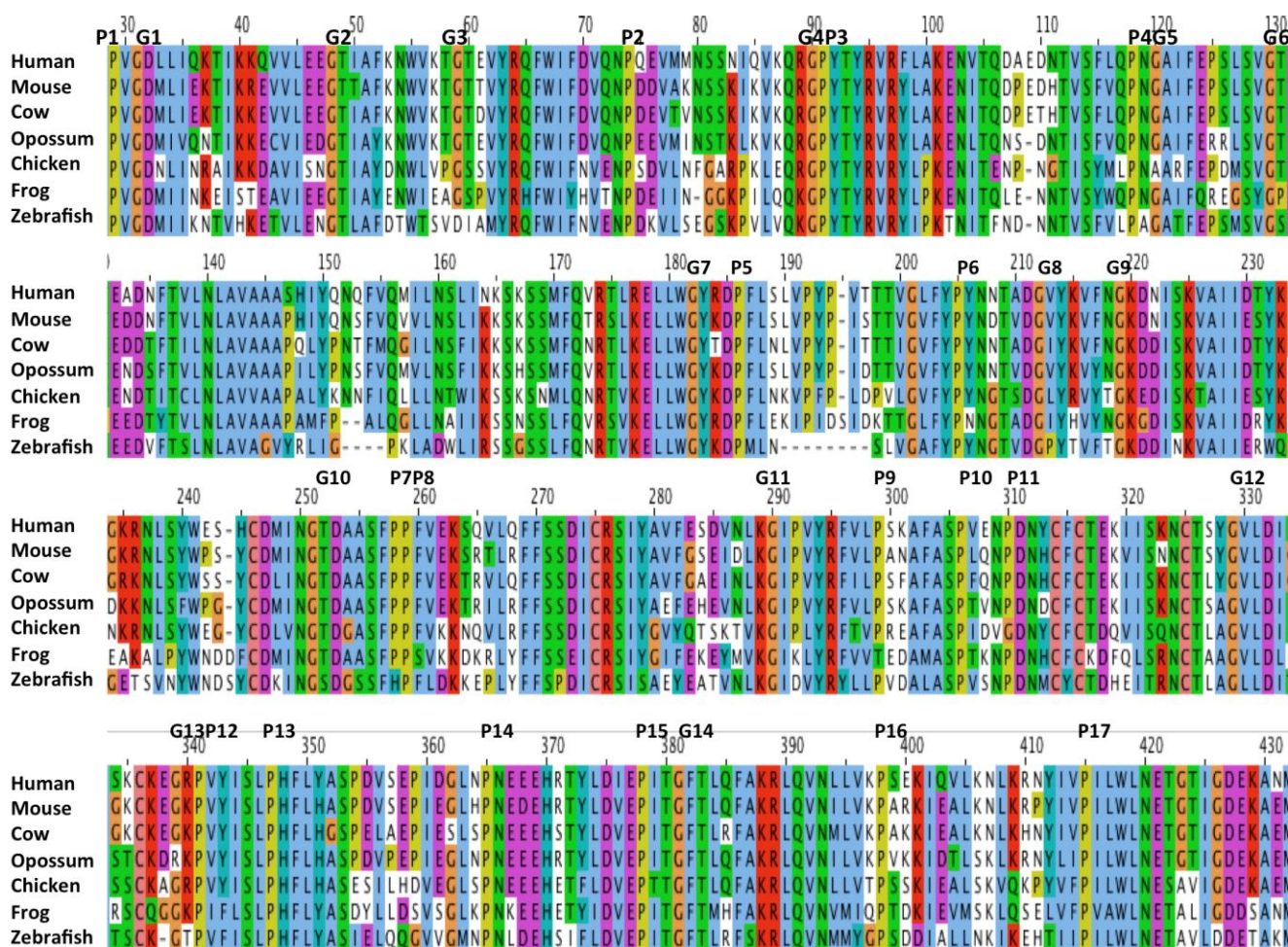
43. Miquilena-Colina, M.E.; Lima-Cabello, E.; Sánchez-Campos, S.; García-Mediavilla, M.V.; Fernández-Bermejo, M.; Lozano-Rodríguez, T.; Vargas-Castrillón, J.; Buqué, X.; Ochoa, B.; Aspichueta, P.; *et al.* Hepatic fatty acid translocase CD36 upregulation is associated with insulin resistance, hyperinsulinaemia and increased steatosis in non-alcoholic steatohepatitis and chronic hepatitis C. *Gut* **2011**, *60*, 1394–1402.
44. Ashraf, M.Z.; Gupta, N. Scavenger receptors: Implications in atherothrombotic disorders. *Int. J. Biochem. Cell Biol.* **2011**, *43*, 697–700.
45. Tandon, N.N.; Lipsky, R.H.; Burgess, W.H.; Jamieson, G.A. Isolation and characterization of platelet glycoprotein IV (CD36). *J. Biol. Chem.* **1989**, *264*, 7570–7575.
46. Wyler, B.; Daviet, L.; Bortkiewicz, H.; Bordet, J.C.; McGregor, J.L. Cloning of the cDNA encoding human platelet CD36: Comparison to PCR amplified fragments of monocyte, endothelial and HEL cells. *Thromb. Haemost.* **1993**, *70*, 500–505.
47. Tao, N.; Wagner, S.J.; Lublin, D.M. CD36 is palmitoylated on both N- and C-terminal cytoplasmic tails. *J. Biol. Chem.* **1996**, *271*, 22315–22320.
48. Chu, L.Y.; Silverstein, R.L. CD36 Ectodomain Phosphorylation Blocks Thrombospondin-1 Binding: Structure-Function Relationships and Regulation by Protein Kinase C. *Arterioscler. Thromb. Vasc. Biol.* **2012**, *32*, 760–767.
49. Rasmussen, J.T.; Berglund, L.; Rasmussen, M.S.; Petersen, T.E. Assignment of disulfide bridges in bovine CD36. *Eur. J. Biochem.* **1998**, *257*, 488–494.
50. Lewandrowski, U.; Moebius, J.; Walter, U.; Sickmann, A. Elucidation of N-glycosylation sites on human platelet proteins: A glycoproteomic approach. *Mol. Cell Proteomics* **2006**, *5*, 226–233.
51. Gupta, R.; Brunak, S. Prediction of glycosylation across the human proteome and the correlation to protein function. *Pac. Symp. Biocomput.* **2002**, *7*, 310–322.
52. Lauzier, B.; Merlen, C.; Vaillant, F.; McDuff, J.; Bouchard, B.; Beguin, P.C.; Dolinsky, V.W.; Foisy, S.; Villeneuve, L.R.; Labarthe, F.; *et al.* Post-translational modifications, a key process in CD36 function: Lessons from the spontaneously hypertensive rat heart. *J. Mol. Cell Cardiol.* **2011**, *51*, 99–108.
53. Gaidukov, L.; Nager, A.R.; Xu, S.; Penman, M.; Krieger, M. Glycine dimerization motif in the N-terminal transmembrane domain of the high density lipoprotein receptor SR-BI required for normal receptor oligomerization and lipid transport. *J. Biol. Chem.* **2011**, *286*, 18452–18464.
54. Salaun, C.; Greaves, J.; Chamberlain, L.H. The intracellular dynamic of protein palmitoylation. *J. Cell Biol.* **2010**, *191*, 1229–1238.
55. Holmes, R.S.; Cox, L.A. Comparative studies of vertebrate scavenger receptor class B type 1 (SCARB1): A high density lipoprotein (HDL) binding protein. *Res. Reports Biochem.* **2012**, *2012*, 9–24.
56. Papale, G.A.; Hanson, P.J.; Sahoo, D. Extracellular disulfide bonds support scavenger receptor class B type I-mediated cholesterol transport. *Biochemistry* **2011**, *50*, 6245–6254.
57. Yua, M.; Romera, K.A.; Nielanda, T.J.F.; Xua, S.; Saenz-Vashb, V.; Penmana, M.; Yesilaltaya, A.; Carrb, S.A.; Kriegera, M. Exoplasmic cysteine Cys384 of the HDL receptor SR-BI is critical for its sensitivity to a small-molecule inhibitor and normal lipid transport activity. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 12243–12248.

58. Kashiwagi, H.; Tomiyama, Y.; Honda, S.; Kosugi, S.; Shiraga, M.; Nagao, N.; Sekiguchi, S.; Kanayama, Y.; Kurata, Y.; Matsuzawa, Y. Molecular basis of CD36 deficiency. Evidence that a 478C-->T substitution (proline90-->serine) in CD36 cDNA accounts for CD36 deficiency. *J. Clin. Invest.* **1995**, *95*, 1040–1046.
59. Yanai, H.; Chiba, H.; Fujiwara, H.; Morimoto, M.; Takahashi, Y.; Hui, S.P.; Fuda, H.; Akita, H.; Kurosawa, T.; Kobayashi, K.; Matsuno, K. Metabolic changes in human CD36 deficiency displayed by glucose loading. *Thromb. Haemost.* **2001**, *86*, 995–999.
60. MacArthur, M.W.; Thornton, J.M. Influence of proline residues on protein conformation. *J. Mol. Biol.* **1991**, *218*, 397–412.
61. Krieger, A.; Möglich, A.; Kiefhaber, T. Effect of proline and glycine residues on dynamics and barriers of loop formation in polypeptide F. *J. Am. Chem. Soc.* **2005**, *127*, 3346–3352.
62. Kay, B.K.; Williamson, M.P.; Sudol, M. The importance of being proline: The interaction of proline-rich motifs in signaling proteins with their cognate domains. *FASEB. J.* **2000**, *14*, 231–241.
63. Coleman, M.D.; Bass, R.B.; Mehan, R.S.; Falke, J.J. Conserved glycine residues in the cytoplasmic domain of the aspartate receptor play essential roles in kinase coupling and on-off switching. *Biochemistry* **2005**, *44*, 7687–7695.
64. Kent, W.J.; Sugnet, C.W.; Furey, T.S. The human genome browser at UCSC. *Genome Res.* **2003**, *12*, 994–1006.
65. Armesilla, A.L.; Calvo, D.; Vega, M.A. Structural and functional characterization of the human CD36 gene promoter: Identification of a proximal PEBP2/CBF site. *J. Biol. Chem.* **1996**, *271*, 7781–7787.
66. Thierry-Mieg, D.; Thierry-Mieg, J. AceView: A comprehensive cDNA-supported gene and transcripts annotation. *Genome Biol.* **2006**, *7*, S12.
67. Sato, O.; Kuriki, C.; Fukui, Y.; Motojima, K. Dual promoter structure of mouse and human fatty acid translocase/CD36 genes and unique transcriptional activation by peroxisome proliferator-activated receptor alpha and gamma ligands. *J. Biol. Chem.* **2002**, *277*, 15703–15711.
68. Olagnier, D.; Lavergne, R.A.; Meunier, E.; Lefèvre, L.; Dardenne, C.; Aubouy, A.; Benoit-Vical, F.; Ryffel, B.; Coste, A.; Berry, A.; *et al.* Nrf2, a PPAR $\gamma$  alternative pathway to promote CD36 expression on inflammatory macrophages: Implication for malaria. *PLoS Pathog.* **2011**, *7*, e1002254.
69. Motojima, K.; Passilly, P.; Peters, J.M.; Gonzalez, F.J.; Latruffe, N. Expression of putative fatty acid transporter genes are regulated by peroxisome proliferator-activated receptor alpha and gamma activators in a tissue- and inducer-specific manner. *J. Biol. Chem.* **1998**, *273*, 16710–16714.
70. Sharrocks, A.D.; von Hesler, F.; Shaw, P.E. The identification of elements determining the different DNA binding specificities of the MADS box proteins p67<sup>SRF</sup> and RSRFC4. *Nucleic Acids Res.* **1993**, *21*, 215–221.
71. Cai, R.L. Human CART1, a paired-class homeodomain protein, activates transcription through palindromic binding sites. *Biochem. Biophys. Res. Commun.* **1998**, *250*, 305–311.
72. Gómez-Ferrer á, M.A.; Rey-Campos, J. Functional domains of FOXJ2. *J Mol Biol.* **2003**, *329*, 631–644.
73. He, Y.; Sun, S.; Sha, H.; Liu, Z.; Yang, L.; Xue, Z.; Chen, H.; Qi, L. Emerging roles for XBP1, a sUPeR transcription factor. *Gene Expr.* **2010**, *15*, 13–25.

74. Dyczkowski, J.; Vingron, M. Comparative analysis of cell cycle regulated genes in eukaryotes. *Genome Inform.* **2005**, *16*, 125–131.
75. Su, A.I.; Wiltshire, T.; Batalov, S.; Lapp, H.; Ching, K.R.; Block, D.; Zhang, J.; Soden, R.; Hayakawa, M.; Kreiman, G.; *et al.* A gene atlas of the human and mouse protein encoding transcriptomes. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 6062–6067.
76. Eguez, L.; Lee, A.; Chavez, J.A.; Miinea, C.P.; Kane, S.; Lienhard, G.E.; McGraw, T.E. Full intracellular retention of GLUT4 requires AS160 Rab GTPase activating protein. *Cell Metab.* **2005**, *2*, 263–272.
77. Thong, F.S.; Bilan, P.J.; Klip, A. The Rab GTPase-activating protein AS160 integrates Akt, protein kinase C, and AMP-activated protein kinase signals regulating GLUT4 traffic. *Diabetes* **2007**, *56*, 414–423.
78. Samovski, D.; Su, X.; Xu, Y.; Abumrad, N.A.; Stahl, P.D. Insulin and AMPK regulate fatty acid translocase/CD36 plasma membrane recruitment in cardiomyocytes via RabGAP AS160 and Rab8a Rab GTPase. *J. Lipid Res.* **2012**, [Epub ahead of print].
79. Sano, H.; Kane, S.; Sano, E.; Miinea, C.P.; Asara, J.M.; Lane, W.S.; Garner, C.W.; Lienhard, G.E. Insulin-stimulated phosphorylation of a Rab GTPase-activating protein regulates GLUT4 translocation. *J. Biol. Chem.* **2003**, *278*, 14599–14602.
80. Yuasa, T.; Uchiyama, K.; Ogura, Y.; Kimura, M.; Teshigawara, K.; Hosaka, T.; Tanaka, Y.; Obata, T.; Sano, H.; Kishi, K.; *et al.* The Rab GTPase-activating protein AS160 as a common regulator of insulin- and Galphaq-mediated intracellular GLUT4 vesicle distribution. *Endocr. J.* **2009**, *56*, 345–359.
81. Liani, R.; Halvorsen, B.; Sestili, S.; Handberg, A.; Santilli, F.; Vazzana, N.; Formoso, G.; Aukrust, P.; Davi, G. Plasma levels of soluble CD36, platelet activation, inflammation, and oxidative stress are increased in type 2 diabetic patients. *Free Radic. Biol. Med.* **2012**, [Epub ahead of print].
82. Mitchell, R.W.; On, N.H.; Del Bigio, M.R.; Miller, D.W.; Hatch, G.M. Fatty acid transport protein expression in human brain and potential role in fatty acid transport across human brain microvessel endothelial cells. *J. Neurochem.* **2011**, *117*, 735–746.
83. Donoghue, P.C.J.; Benton, M.J. Rocks and clocks: Calibrating the tree of life using fossils and molecules. *Trends Genet.* **2007**, *22*, 424–431.
84. Bentley, A.A.; Adams, J.C. The evolution of thrombospondins and their ligand-binding activities. *Mol. Biol. Evol.* **2010**, *27*, 2187–2197.
85. Altschul, F.; Vyas, V.; Cornfield, A.; Goodin, S.; Ravikumar, T.S.; Rubin, E.H.; Gupta, E. Basic local alignment search tool. *J. Mol. Biol.* **1997**, *215*, 403–410.
86. McGuffin, L.J.; Bryson, K.; Jones, D.T. The PSIPRED protein structure prediction server. *Bioinform.* **2000**, *16*, 404–405.
87. Hall, T.A. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **1999**, *41*, 95–99.
88. Felsenstein, J. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* **1986**, *39*, 783–789.



**Supplementary Figure 1.** Amino Acid Sequence Alignments for Vertebrate CD36 Exoplasmic Sequences. Amino acids are color coded: yellow for proline (P); S (serine); green for hydrophilic amino acids, S (serine), Q (glutamine), N (asparagine), and T (threonine); brown for glycine (G); light blue for hydrophobic amino acids, L (leucine), I (isoleucine), V (valine), M (methionine), W (tryptophan); dark blue for amino acids, T (tyrosine) and H (histidine); purple for acidic amino acids, E (glutamate) and D (aspartate); and red for basic amino acids, K (lysine) and R (arginine); conserved prolines and glycines are designated as P1, P2 etc and G1, G2 etc, respectively. Numbers refer to human CD36 amino acid sequence.



**Supplementary Table 1.** CD36, SCARB1 and SCARB2 proteins: subunit MWs and percentage identities. High % identities are in **bold**.

CD36 Gene	Species	Subunit MW	% Identity with human	% Identity with human	% Identity with human
			SCARB1	SCARB2	CD36
Human	<i>Homo sapiens</i>	53,053	31	30	<b>100</b>
Chimpanzee	<i>Pan troglodytes</i>	53,064	31	30	<b>100</b>
Orangutan	<i>Pongo abelii</i>	53,039	32	30	<b>97</b>
Gibbon	<i>Nomascus leucogenys</i>	53,161	32	30	<b>96</b>
Rhesus	<i>Macaca mulatta</i>	53,041	32	31	<b>94</b>
Mouse	<i>Mus musculus</i>	52,698	30	31	<b>83</b>
Rat	<i>Rattus norvegicus</i>	52,731	31	30	<b>86</b>
Guinea Pig	<i>Cavia porcellus</i>	53,085	32	32	<b>81</b>
Cow	<i>Bos taurus</i>	52,940	32	30	<b>82</b>
Dog	<i>Canis familiaris</i>	52,549	31	30	<b>82</b>
Pig	<i>Sus scrofa</i>	53,085	31	30	<b>82</b>
Rabbit	<i>Oryctolagus cuniculus</i>	52,729	31	31	<b>88</b>
Horse	<i>Equus caballus</i>	52,789	31	31	<b>83</b>
Elephant	<i>Loxodonta africana</i>	52,873	31	31	<b>80</b>
Opossum	<i>Monodelphis domestica</i>	53,017	30	30	<b>73</b>
Platypus	<i>Ornithorhynchus anatinus</i>	52,807	31	30	<b>73</b>
Chicken	<i>Gallus gallus</i>	52,624	30	32	<b>61</b>
Lizard	<i>Anolis carolinensis</i>	52,890	31	31	<b>61</b>
Frog	<i>Xenopus tropicalis</i>	52,696	30	29	<b>55</b>
Zebrafish	<i>Danio rerio</i>	51,590	31	31	<b>53</b>

SCARB1 Gene	Species	Subunit MW	% Identity with human	% Identity with human	% Identity with human
			SCARB1	SCARB2	CD36
Human	<i>Homo sapiens</i>	56,973	<b>100</b>	29	31
Mouse	<i>Mus musculus</i>	56,754	<b>79</b>	29	29
Chicken	<i>Gallus gallus</i>	55,918	<b>57</b>	28	31
Zebrafish	<i>Danio rerio</i>	55,742	<b>51</b>	28	30
<b>SCARB2 Gene</b>					
Human	<i>Homo sapiens</i>	54,290	29	<b>100</b>	30
Mouse	<i>Mus musculus</i>	54,044	29	<b>85</b>	31
Chicken	<i>Gallus gallus</i>	53,907	30	<b>59</b>	33
Zebrafish	<i>Danio rerio</i>	60,234	31	<b>43</b>	33
<b>CD36 Gene</b>					
Lancelet	<i>Branchiostoma floridae</i>	54,141	34	35	35
Sea squirt	<i>Ciona intestinalis</i>	58,009	26	33	31
Nematode	<i>Caenorhabditis elegans</i>	60,182	21	26	24
Fruit fly	<i>Drosophila melanogaster</i>	58,663	20	23	26