Scavenger receptor class B member 2 (SCARB2) (also called LIMP-2, CD36L2 or LGP85) is a major lysosomal membrane glycoprotein involved in endosomal and lysosomal biogenesis and maintenance. SCARB2 acts as a receptor for the lysosomal mannose-6-phosphate independent targeting of β-glucuronidase and enterovirus 71 and influences Parkinson's disease and epilepsy. Genetic deficiency of this protein causes deafness and peripheral neuropathy in mice as well as myoclonic epilepsy and nephrotic syndrome in humans. Comparative SCARB2 amino acid sequences and structures and SCARB2 gene locations were examined using data from several vertebrate genome projects. Vertebrate SCARB2 sequences shared 43-100% identity as compared with 30-36% sequence identities with other CD36-like superfamily members, SCARB1 and CD36. At least 10 N-glycosylation sites were conserved among most vertebrate SCARB2 proteins examined. Sequence alignments, key amino acid residues and conserved predicted secondary structures were examined, including cytoplasmic, transmembrane and external lysosomal membrane sequences: cysteine disulfide residues, thrombospondin (THP1) binding sites and 16 proline and 20 glycine conserved residues, which may contribute to short loop formation within the exomembrane SCARB2 sequences. Vertebrate SCARB2 genes contained 12 coding exons. The human SCARB2 gene contained a CpG island (CpG100), ten microRNA-binding sites and several transcription factor binding sites (including PPARA) which may contribute to a higher level (2.4 times average) of gene expression. Phylogenetic analyses examined the relationships and potential evolutionary origins of the vertebrate SCARB2 gene with vertebrate SCARB1 and CD36 genes. These suggested that SCARB2 originated from duplications of the CD36 gene in an ancestral genome forming three vertebrate CD36 gene family members: SCARB1, SCARB2 and CD36.
recognized as a candidate gene for autism (Ilu et al. 2010) and Parkinson’s disease (Do et al. 2011, Michelakakis et al. 2012).

SCARB1 (also called CLA1, SRB1 and CD36L1) is a second member of the CD36 family that serves as a homo-oligomeric plasma membrane cell surface glycoprotein receptor for high density lipoprotein cholesterol (HDL), other phospholipid ligands and chylomicron remnants (Acton et al. 1996, Bulte-Brienne et al. 2002, Connelly et al. 2004, Holmes and Cox 2012, Kent & Stylianou 2011, Marsche et al. 2003). A third member of the CD36 family, CD36 (also called SCARB3, fatty acyltranslocase [FAT] and glycoprotein 88 [GP88]) 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 low density lipoprotein cholesterol (LDL) (Martin et al. 2007, Tandon et al. 1989), and has been implicated in several diseases including insulin resistance, diabetes, atherosclerosis and malaria (Adachi & Tsujimoto 2006, Collot-Teixeira et al. 2007, Gautum & Banerjee 2011, Martin et al. 2007, Ren 2012, Simantov & Silverstein 2003).

The gene encoding human SCARB2 (SCARB2) is on chromosome 4, encoded by 12 coding exons (Calvo et al. 1995) and localized between the genes encoding nucleoporin (NUP54) and protein family 47 (FAM47E) (Kent et al. 2003). In addition, SCARB2 has been assigned to chromosome 8 in pigs (Kim et al. 2006) and to chromosome 5 in mice, but designated as Scarb2 in the latter genome (Tabuchi et al. 1997). SCARB2 is ubiquitously expressed in various cells and tissues of the body, including kidney glomerular tubules (Berkovic et al. 2008, Desmond et al. 2011), liver (Tabuchi et al. 1997, Zhang et al. 2007), retinal ganglia and photoreceptor outer segment segments (Tserentsoodol et al. 2006), ureter epithelial cells (Gamp et al. 2003), metastatic pancreas islet cells (Fujita et al. 1992), the cardiac intercalated disc (Schroen et al. 2007) and phagocytes (Carrasco-Marín et al. 2011). Studies of Scarcb2/Scarb2−/− knock out mice have shown that SCARB2-deficiency causes ureteric pelvic junction obstruction, deafness and peripheral neuropathy (Gamp et al. 2003), renal tubular proteinuria (Desmond et al. 2011), an inability to mount a hypertrophic response to increased blood pressure, due to the absence of SCARB2 in the cardiac intercalated disc (Schroen et al. 2007) and macrophage-related defects in immunity to infection (Carrasco-Marín et al. 2011). Human clinical studies have also examined SCARB2 polymorphisms associated with diseases causing mutations of β-glucocerebrosidase binding, which is defective in Gaucher disease (Blanz et al. 2010, Velayati et al. 2011); and collapsing focal and segmental glomerular sclerosis (FSGS) and myoclonic epilepsy (Balreira et al. 2008, Berkovic et al. 2008, Chaves et al. 2011, Desmond et al. 2011, Dibbens et al. 2011, Hopfner et al. 2011, Rubboli et al. 2011).

This paper reports the predicted gene structures and amino acid sequences for several vertebrate SCARB2 genes and proteins, the secondary structures for vertebrate SCARB2 proteins, several potential sites for regulating human SCARB2 gene expression and the structural, phylogenetic and evolutionary relationships for these genes and enzymes with those for vertebrate SCARB2, SCARB1 and CD36 gene families.

Materials and Methods

Vertebrate SCARB2 gene and protein identification
BLAST (Basic Local Alignment Search Tool) studies were undertaken using web tools from the National Center for Biotechnology Information (NCBI) (http://blast.ncbi.nlm.nih.gov/Blast.cgi) (Altschul et al. 1997). Protein BLAST analyses used human and mouse SCARB2 amino acid sequences previously described (Fujita et al. 1992, Ogata et al. 1994, Calvo et al. 1995, Tabuchi et al. 1997) (Table 1). Non-redundant protein sequence databases for several vertebrate genomes were examined using the blastp algorithm from sources previously described (Holmes 2012). 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 SCARB2-like proteins. Predicted SCARB2-like protein sequences were obtained in each case and subjected to analyses of predicted protein and gene structures.

BLAT (Blast-like Alignment Tool) analyses were subsequently undertaken for each of the predicted SCARB2 amino acid sequences using the UC Santa Cruz Genome Browser (Kent et al. 2003) with the default settings to obtain the predicted locations for each of the vertebrate SCARB2 genes, including predicted exon boundary locations and gene sizes. BLAT analyses were similarly undertaken for vertebrate SCARB1 and CD36 genes using previously reported sequences in each case (see Table 1; Holmes & Cox 2012). Structures for human SCARB2, mouse Scarb2 and rat Scarb2 transcripts were obtained using the AceView website to examine predicted gene and protein structures (Thierry-Mieg & Thierry-Mieg 2006). Predictions for human, mouse and rat SCARB2 CpG islands, miRNA binding sites and transcription factor binding sites were obtained using the UC Santa Cruz Genome Browser (Kent et al. 2003).
<table>
<thead>
<tr>
<th>SCARB2 Gene</th>
<th>Species</th>
<th>RefSeq ID</th>
<th>GenBank ID</th>
<th>UNIPROT ID</th>
<th>Amino acids</th>
<th>Chromosome location</th>
<th>Coding Exons (strand)</th>
<th>Gene Size bps</th>
<th>Subunit MW</th>
<th>Gene Expression Level</th>
<th>% Identity with human SCARB1</th>
<th>% Identity with human SCARB2</th>
<th>% Identity with human CD36</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>Homo sapiens</td>
<td>NM_005506</td>
<td>BT006939</td>
<td>Q53Y63</td>
<td>478</td>
<td>4:77,084,378-77,134,696</td>
<td>12 (-ve)</td>
<td>50,316</td>
<td>54,290</td>
<td>3.2</td>
<td>34</td>
<td>100</td>
<td>35</td>
</tr>
<tr>
<td>Chimpanzee</td>
<td>Pan troglodytes</td>
<td>XP_517214.2</td>
<td>na</td>
<td>na</td>
<td>478</td>
<td>4:77,134,567-77,135,377</td>
<td>12 (+ve)</td>
<td>50,633</td>
<td>54,309</td>
<td>na</td>
<td>34</td>
<td>100</td>
<td>35</td>
</tr>
<tr>
<td>Orangutan</td>
<td>Pongo abelii</td>
<td>XP_02814942.1</td>
<td>na</td>
<td>na</td>
<td>478</td>
<td>4:79,503,577-79,555,051</td>
<td>12 (-ve)</td>
<td>51,475</td>
<td>54,331</td>
<td>na</td>
<td>33</td>
<td>99</td>
<td>34</td>
</tr>
<tr>
<td>Marmoset</td>
<td>Callithrix jacchus</td>
<td>XP_002745738.1</td>
<td>na</td>
<td>na</td>
<td>478</td>
<td>4:318,471,528-118,533,432</td>
<td>12 (+ve)</td>
<td>61,905</td>
<td>54,215</td>
<td>na</td>
<td>32</td>
<td>97</td>
<td>34</td>
</tr>
<tr>
<td>Mouse</td>
<td>Mus musculus</td>
<td>NM_007644</td>
<td>BC029073</td>
<td>O35114</td>
<td>478</td>
<td>5:92,875,334-92,934,732</td>
<td>12 (+ve)</td>
<td>61,905</td>
<td>54,215</td>
<td>na</td>
<td>32</td>
<td>97</td>
<td>34</td>
</tr>
<tr>
<td>Rat</td>
<td>Rattus norvegicus</td>
<td>NM_054001</td>
<td>BC061853</td>
<td>P27615</td>
<td>478</td>
<td>14:17,119,648-17,168,942</td>
<td>12 (+ve)</td>
<td>51,485</td>
<td>54,091</td>
<td>1.0</td>
<td>32</td>
<td>86</td>
<td>36</td>
</tr>
<tr>
<td>Panda</td>
<td>Ailuropoda melanoleuca</td>
<td>XP_002924780.1</td>
<td>na</td>
<td>na</td>
<td>478</td>
<td>*GL191385.1-272,540-348,835</td>
<td>12 (+ve)</td>
<td>76,296</td>
<td>54,059</td>
<td>na</td>
<td>32</td>
<td>87</td>
<td>34</td>
</tr>
<tr>
<td>Cow</td>
<td>Bos taurus</td>
<td>NM_00110213.1</td>
<td>BC149935</td>
<td>A6QQP4</td>
<td>478</td>
<td>6:92,797,818-92,879,565</td>
<td>12 (-ve)</td>
<td>81,748</td>
<td>53,985</td>
<td>3.6</td>
<td>32</td>
<td>85</td>
<td>35</td>
</tr>
<tr>
<td>Dog</td>
<td>Canis familiaris</td>
<td>XP_535612.4</td>
<td>na</td>
<td>F1PGJ5</td>
<td>478</td>
<td>3:32,756,732-3,808,324</td>
<td>12 (+ve)</td>
<td>51,593</td>
<td>54,015</td>
<td>na</td>
<td>31</td>
<td>86</td>
<td>34</td>
</tr>
<tr>
<td>Pig</td>
<td>Sus scrofa</td>
<td>NM_00124455.1</td>
<td>na</td>
<td>F1RYT3</td>
<td>478</td>
<td>8:61,286,560-61,335,361</td>
<td>12 (+ve)</td>
<td>66,802</td>
<td>54,015</td>
<td>na</td>
<td>34</td>
<td>88</td>
<td>35</td>
</tr>
<tr>
<td>Rabbit</td>
<td>Oryctolagus cuniculus</td>
<td>XP_002717150.1</td>
<td>na</td>
<td>G1SJL6</td>
<td>478</td>
<td>15:574169938-7423068</td>
<td>12 (+ve)</td>
<td>63,131</td>
<td>53,588</td>
<td>na</td>
<td>34</td>
<td>90</td>
<td>33</td>
</tr>
<tr>
<td>Elephant</td>
<td>Loxodonta africana</td>
<td>XP_003414189.1</td>
<td>na</td>
<td>G3T2P7</td>
<td>478</td>
<td>15:30,304,405-21,106,415</td>
<td>12 (+ve)</td>
<td>76,011</td>
<td>53,981</td>
<td>na</td>
<td>33</td>
<td>87</td>
<td>35</td>
</tr>
<tr>
<td>Chicken</td>
<td>Gallus gallus</td>
<td>XP_42093.1</td>
<td>B931548</td>
<td>na</td>
<td>481</td>
<td>4:45,11,268-146,901,887</td>
<td>12 (+ve)</td>
<td>18,353</td>
<td>53,907</td>
<td>na</td>
<td>30</td>
<td>59</td>
<td>33</td>
</tr>
<tr>
<td>Lizard</td>
<td>Anolis carolinensis</td>
<td>XP_003221897.1</td>
<td>na</td>
<td>G1KMI7</td>
<td>482</td>
<td>5:514,878,187-514,878,187</td>
<td>12 (+ve)</td>
<td>23,701</td>
<td>53,969</td>
<td>na</td>
<td>33</td>
<td>53</td>
<td>33</td>
</tr>
<tr>
<td>Frog</td>
<td>Xenopus tropicalis</td>
<td>NM_001016577.2</td>
<td>BC171025</td>
<td>B7ZTX2</td>
<td>483</td>
<td>146,901,887-63,699</td>
<td>12 (+ve)</td>
<td>32,825</td>
<td>54,586</td>
<td>na</td>
<td>31</td>
<td>54</td>
<td>35</td>
</tr>
<tr>
<td>Zebrafish</td>
<td>Danio rerio</td>
<td>NM_173259.1</td>
<td>BC162407</td>
<td>Q8JQR8</td>
<td>474</td>
<td>5:63,942,096-63,955,449</td>
<td>13 (+ve)</td>
<td>13,354</td>
<td>60,234</td>
<td>na</td>
<td>31</td>
<td>43</td>
<td>33</td>
</tr>
<tr>
<td>SCARB1 Gene</td>
<td>Species</td>
<td>RefSeq ID &quot;Ensembl/NCBI&quot;</td>
<td>GenBank ID</td>
<td>UNIPROT ID</td>
<td>Amino acids</td>
<td>Chromosome location</td>
<td>Coding Exons (strand)</td>
<td>Gene Size bps</td>
<td>Subunit MW</td>
<td>Gene Expression Level</td>
<td>% Identity with human SCARB1</td>
<td>% Identity with human SCARB2</td>
<td>% Identity with human CD36</td>
</tr>
<tr>
<td>-------------</td>
<td>-------------------------</td>
<td>--------------------------</td>
<td>------------</td>
<td>------------</td>
<td>-------------</td>
<td>---------------------</td>
<td>----------------------</td>
<td>---------------</td>
<td>------------</td>
<td>-----------------------</td>
<td>----------------------------</td>
<td>-----------------------------</td>
<td>---------------------------</td>
</tr>
<tr>
<td>Human</td>
<td>Homo sapiens</td>
<td>NM_00505</td>
<td>BC020087</td>
<td>Q8WVT0</td>
<td>509</td>
<td>12:125,267,232-125,348,266</td>
<td>12 (-ve)</td>
<td>81,035</td>
<td>56,973</td>
<td>13.7</td>
<td>100</td>
<td>29</td>
<td>31</td>
</tr>
<tr>
<td>Mouse</td>
<td>Mus musculus</td>
<td>NM_001205082.1</td>
<td>BC004656</td>
<td>Q61009</td>
<td>509</td>
<td>5:125,761,478-125,821,252</td>
<td>12 (-ve)</td>
<td>63,985</td>
<td>56,754</td>
<td>5.1</td>
<td>79</td>
<td>29</td>
<td>29</td>
</tr>
<tr>
<td>Chicken</td>
<td>Gallus gallus</td>
<td>'XP_415106'</td>
<td>na</td>
<td>na</td>
<td>503</td>
<td>15:4,543,054-4,558,954</td>
<td>12 (+ve)</td>
<td>15,901</td>
<td>55,918</td>
<td>na</td>
<td>57</td>
<td>28</td>
<td>31</td>
</tr>
<tr>
<td>Zebrafish</td>
<td>Danio rerio</td>
<td>NM_198121</td>
<td>BC044516</td>
<td>E7FB50</td>
<td>496</td>
<td>11:21,526,513-21,572,478</td>
<td>12 (-ve)</td>
<td>45,684</td>
<td>55,742</td>
<td>na</td>
<td>51</td>
<td>28</td>
<td>30</td>
</tr>
<tr>
<td>CD36 Gene</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human</td>
<td>Homo sapiens</td>
<td>NM_001001547</td>
<td>BC008406</td>
<td>P16671</td>
<td>472</td>
<td>7:80,275,645-80,303,732</td>
<td>12 (+ve)</td>
<td>72,231</td>
<td>53,053</td>
<td>6.6</td>
<td>31</td>
<td>30</td>
<td>100</td>
</tr>
<tr>
<td>Mouse</td>
<td>Mus musculus</td>
<td>NM_001159555.1</td>
<td>BC010262</td>
<td>Q08857</td>
<td>472</td>
<td>5:17,291,543-17,334,712</td>
<td>12 (-ve)</td>
<td>43,170</td>
<td>52,698</td>
<td>4.2</td>
<td>30</td>
<td>31</td>
<td>83</td>
</tr>
<tr>
<td>Chicken</td>
<td>Gallus gallus</td>
<td>'ENSGALG8439'</td>
<td>AJ719746</td>
<td>F1NER9</td>
<td>471</td>
<td>1:12,077,308-12,107,415</td>
<td>12 (-ve)</td>
<td>30,108</td>
<td>52,624</td>
<td>na</td>
<td>30</td>
<td>32</td>
<td>61</td>
</tr>
<tr>
<td>Zebrafish</td>
<td>Danio rerio</td>
<td>NP_001002363.1</td>
<td>BC076048</td>
<td>Q6DHC7</td>
<td>465</td>
<td>4:21,594,449-21,606,961</td>
<td>12 (-ve)</td>
<td>12,513</td>
<td>51,590</td>
<td>na</td>
<td>31</td>
<td>31</td>
<td>53</td>
</tr>
<tr>
<td>Lancelet</td>
<td>Branchiostoma floridae</td>
<td>'XP_002609178.1'</td>
<td>na</td>
<td>na</td>
<td>480</td>
<td>Un:534,334,234-534,343,082</td>
<td>12 (+ve)</td>
<td>8,849</td>
<td>54,144</td>
<td>na</td>
<td>34</td>
<td>35</td>
<td>35</td>
</tr>
<tr>
<td>Sea squirt</td>
<td>Ciona intestinalis</td>
<td>'XP_002127015.1'</td>
<td>na</td>
<td>na</td>
<td>523</td>
<td>09p:2,872,362-2,873,903</td>
<td>1 (-ve)</td>
<td>1,542</td>
<td>58,009</td>
<td>na</td>
<td>26</td>
<td>33</td>
<td>31</td>
</tr>
<tr>
<td>Nematode</td>
<td>Caenorhabditis elegans</td>
<td>NM_067224</td>
<td>na</td>
<td>Q9XTT3</td>
<td>534</td>
<td>III:12,453,609-12,456,726</td>
<td>8 (+ve)</td>
<td>3,118</td>
<td>60,182</td>
<td>4.6</td>
<td>21</td>
<td>26</td>
<td>24</td>
</tr>
<tr>
<td>Fruit fly</td>
<td>Drosophila melanogaster</td>
<td>NP_523859</td>
<td>na</td>
<td>na</td>
<td>520</td>
<td>2R:20,864,606-20,867,116</td>
<td>6 (-ve)</td>
<td>#2,511</td>
<td>58,663</td>
<td>na</td>
<td>20</td>
<td>23</td>
<td>26</td>
</tr>
</tbody>
</table>
Predicted Structures and Properties of Vertebrate SCARB2 and other CD36-like Proteins

Predicted secondary structures for vertebrate SCARB2 proteins, human SCARB1 and CD36, 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 (McGuffin et al. 2000). Molecular weights, N-glycosylation sites (Gupta & Brunak 2002) and predicted transmembrane, cytosolic and extracellular sequences for vertebrate SCARB2 proteins were obtained using Expasy web tools (http://au.expasy.org/tools/pi_tool.html).

Comparative Human SCARB2 and Mouse Scarb2 Gene Expression

The genome browser (http://genome.ucsc.edu) (Kent et al. 2003) was used to examine GNF Expression Atlas 2 data using various expression chips for human SCARB2 and mouse Scarb2 genes, respectively (Su et al. 2004) (http://biogps.gnf.org). Gene array expression ‘heat maps’ were examined for comparative gene expression levels among human and mouse tissues showing high (red), intermediate (black), and low (green) expression levels.

Phylogeny Studies and Sequence Divergence

Alignments of vertebrate SCARB2, SCARB1 and CD36 sequences were assembled using BioEdit v.5.0.1 using the default settings (Hall 1999). Alignment of ambiguous regions, including the amino and carboxyl termini, were excluded prior to phylogenetic analysis, yielding alignments of 431 residues for comparisons of vertebrate SCARB2 sequences with human, mouse, chicken and zebrafish SCARB1 and CD36 sequences with the lancelet (Branchiostoma floridae) CD36 sequence (Table 1). Evolutionary distances and phylogenetic trees were calculated as previously described (Holmes 2012). Tree topology was reexamined by the boot-strap method (100 bootstraps were applied) of resampling and only values that were highly significant (≥95) are shown (Felsenstein 1985).

Results and Discussion

Alignments of Vertebrate SCARB2 Amino Acid Sequences

The deduced amino acid sequences for cow (Co) (Bos taurus), chicken (Ch) (Gallus gallus), lizard (Li) (Anolis carolensis), frog (Fr) (Xenopus tropicalis) and zebrafish (Zf) (Danio rerio) SCARB2 genes along with predicted secondary structures and properties were obtained as described above.

Figure 1. Amino Acid Sequence Alignments for Vertebrate SCARB2 Sequences. See Table 1 for sources of SCARB2 sequences; * shows identical residues for SCARB2 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 highlighted in green; predicted disulfide bond Cys residues are highlighted in cyan blue; predicted α-helices for vertebrate SCARB2 are highlighted in yellow and numbered in sequence from the start of the predicted exoplasmic domain; predicted β-sheets are highlighted in grey and also numbered in sequence; bold underlined font shows residues corresponding to known or predicted exon start sites; exon numbers refer to human SCARB2 gene exons; exolsosomic refers to the predicted SCARB2 sequence external to the lysosomal membrane.
zebrafish (Zf) (Danio rerio) SCARB2 are shown in Figure 1 together with the previously reported sequences for human (Hu) and mouse (Mo) SCARB2 (Table 1) (Fujita et al. 1992, Tabuchi et al. 1997). Alignments of human with other vertebrate SCARB2 sequences examined were between 43-100% identical, suggesting that these are products of the same family of genes, whereas comparisons of sequence identities of vertebrate SCARB2 proteins with human SCARB1 and CD36 proteins exhibited lower levels of sequence identities (30-36%), indicating that these are members of distinct CD36-like gene families (Table 1). The amino acid sequences for mammalian SCARB2 contained 478 residues while chicken (Gallus gallus), lizard (Anolis carolensis), frog (Xenopus tropicalis) and zebrafish (Danio rerio) SCARB2 contained 481, 482, 483 and 474 amino acids, respectively (Table 1; Figure 1).

Previous studies have reported several key regions and residues for human and mouse SCARB2 proteins (human SCARB2 amino acid residues were identified in each case). These included cytoplasmic N-terminal and C-terminal residues, residues 2-6 and 457-478, and N-terminal and C-terminal transmembrane helical regions: residues 7-30 and 432-456 (Fujita et al. 1992, Tabuchi et al. 1997). These motifs underwent significant changes in amino acid sequence but retained the predicted cytoplasmic and transmembrane properties in each case (Figure 1). These changes are in contrast to the N-terminal transmembrane sequences for vertebrate SCARB1 and CD36 sequences, for which several glycine residues were predominantly conserved, especially for CD36 Gly12, Gly16 and Gly24/Gly25 residues and for SCARB1 key N-terminal glycine residues (Gly15/Gly18/Gly25) (Holmes & Cox 2012), which form a dimerization motif in the N-terminal transmembrane domain and participate in forming SCARB1 oligomers (Gaidukov et al. 2011). There were no SCARB2 N-terminal transmembrane glycine residues conserved for the vertebrate sequences examined, although Gly10 was predominantly conserved with a Gly/Ala substitution observed for the frog SCARB2 sequence (Figure 1). A conserved glycine residue was observed for the vertebrate C-terminal transmembrane sequences (human SCARB2 Gly449) (Figure 1); however the role of this residue has not been investigated.

Comparative Sequences for Vertebrate SCARB2 N-Glycosylation Sites

Ten N-glycosylation sites for human SCARB2 have been previously identified for this protein (Figure 1; Table 2) (Lewandrowski et al. 2005), whereas eleven such sites have been reported for mouse SCARB2 (Tabuchi et al. 1997). All of these sites were predominantly retained among the 16 vertebrate SCARB2 sequences examined. Given the sequence conservation observed for these residues among the vertebrate SCARB2 sequences examined, it is apparent that they are essential for the function of vertebrate SCARB2 as a glycoprotein. The multiple N-glycosylation sites observed for vertebrate SCARB2 sequences were consistent with a major role for N-proteoglycan residues exposed on the external membrane surface of lysosomes in the performance of SCARB2 functions in binding various lipid molecules, and in their reported functions in maintaining the organization of lysosomal and endosomal membranes (Gamp et al. 2003; Kuronita et al. 2002). This is also supported by recent animal model studies, which demonstrated a key role for N-glycosylation in the recruitment of a related integrated membrane CD36-like family member (CD36) into cardiac membranes (Lauzier et al. 2011).

Vertebrate SCARB2 Cysteine Residues

Four conserved external lysosomal membrane vertebrate SCARB2 cysteine residues were observed: Cys274, Cys312, Cys318 and Cys329, which corresponded to four of six previously identified disulfide forming cysteine residues for bovine CD36 (Rasmussen et al. 1998). In contrast, 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) (Holmes & Cox 2012). The CD36 N- and C-terminal conserved cytoplasmic cysteine residues have been shown to be palmitoylated (Tao et al. 1996), which may contribute to protein-protein interactions, protein trafficking and membrane localization (Salau et al. 2010). These conserved cysteines are lacking in the vertebrate SCARB2 sequences (Figure 1), which suggests that S-palmitoyl cysteine residues do not play a role for this lysosomal membrane protein. Vertebrate SCARB1 exoplasmic sequences also contained only four conserved cysteine residues forming disulfide bridges (Cys281; Cys321; Cys323; and Cys334) (Holmes & Cox 2012) although another conserved SCARB1 cysteine (not observed in the CD36 and SCARB2 sequences) (human SCARB1 Cys384) serves a major role in lipid transfer activity (Papale et al. 2011, Yua et al. 2011) (Figure 2).

SCARB2 C-terminal Lysosomal Targeting Sequences

The targeting of SCARB2 to lysosomes has been previously shown to result from a Leu-Ile dipeptide motif
Table 2. Predicted N-glycosylation Sites for Vertebrate SCARB2 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 12 potential sites identified, including 10 sites for human SCARB2. 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 bold.

<table>
<thead>
<tr>
<th>Vertebrate SCARB2</th>
<th>Species</th>
<th>Site 1</th>
<th>Site 2</th>
<th>Site 3</th>
<th>Site 4</th>
<th>Site 5</th>
<th>Site 6</th>
<th>Site 7</th>
<th>Site 8</th>
<th>Site 9</th>
<th>Site 10</th>
<th>Site 11</th>
<th>Site 12</th>
<th>Site 13</th>
<th>Site 14</th>
<th>No of Sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>Homo sapiens</td>
<td>45</td>
<td>NGTE</td>
<td>68</td>
<td>NVTN</td>
<td>105</td>
<td>NGTT</td>
<td>206</td>
<td>NGTN</td>
<td>224</td>
<td>NFTK</td>
<td>249</td>
<td>NGTD</td>
<td>304</td>
<td>NTSI</td>
<td>325</td>
</tr>
<tr>
<td>Chimp</td>
<td>Pan troglodytes</td>
<td>45</td>
<td>NGTE</td>
<td>68</td>
<td>NVTN</td>
<td>105</td>
<td>NGTT</td>
<td>206</td>
<td>NGTN</td>
<td>224</td>
<td>NFTK</td>
<td>249</td>
<td>NGTD</td>
<td>304</td>
<td>NTSI</td>
<td>325</td>
</tr>
<tr>
<td>Orangutan</td>
<td>Pongo abelii</td>
<td>45</td>
<td>NGTE</td>
<td>68</td>
<td>NVTN</td>
<td>105</td>
<td>NGTT</td>
<td>206</td>
<td>NGTN</td>
<td>224</td>
<td>NFTK</td>
<td>249</td>
<td>NGTD</td>
<td>304</td>
<td>NTSI</td>
<td>325</td>
</tr>
<tr>
<td>Rhesus</td>
<td>Macaca mulatta</td>
<td>45</td>
<td>NGTE</td>
<td>68</td>
<td>NVTN</td>
<td>105</td>
<td>NGTT</td>
<td>206</td>
<td>NGTN</td>
<td>224</td>
<td>NFTK</td>
<td>249</td>
<td>NGTD</td>
<td>304</td>
<td>NTSI</td>
<td>325</td>
</tr>
<tr>
<td>Marmoset</td>
<td>Callithrix jacchus</td>
<td>45</td>
<td>NGTE</td>
<td>68</td>
<td>NVTN</td>
<td>105</td>
<td>NGTT</td>
<td>206</td>
<td>NGTN</td>
<td>224</td>
<td>NFTK</td>
<td>249</td>
<td>NGTD</td>
<td>304</td>
<td>NTSI</td>
<td>325</td>
</tr>
<tr>
<td>Mouse</td>
<td>Mus musculus</td>
<td>45</td>
<td>NGTK</td>
<td>68</td>
<td>NVTN</td>
<td>105</td>
<td>NGTT</td>
<td>122</td>
<td>NQSV</td>
<td>206</td>
<td>NGTN</td>
<td>224</td>
<td>NFTK</td>
<td>249</td>
<td>NGTD</td>
<td>304</td>
</tr>
<tr>
<td>Rat</td>
<td>Rattus norvegicus</td>
<td>45</td>
<td>NGTK</td>
<td>68</td>
<td>NVTN</td>
<td>105</td>
<td>NGTT</td>
<td>122</td>
<td>NQSV</td>
<td>206</td>
<td>NGTN</td>
<td>224</td>
<td>NFTK</td>
<td>249</td>
<td>NGTD</td>
<td>304</td>
</tr>
<tr>
<td>Cow</td>
<td>Bos taurus</td>
<td>45</td>
<td>NQSE</td>
<td>68</td>
<td>NVTN</td>
<td>105</td>
<td>NGTT</td>
<td>206</td>
<td>NGTN</td>
<td>224</td>
<td>NFTK</td>
<td>249</td>
<td>NGTD</td>
<td>304</td>
<td>NTSI</td>
<td>325</td>
</tr>
<tr>
<td>Dog</td>
<td>Canis familiaris</td>
<td>45</td>
<td>NQSE</td>
<td>68</td>
<td>NVTN</td>
<td>105</td>
<td>NGTT</td>
<td>122</td>
<td>NQSV</td>
<td>206</td>
<td>NGTN</td>
<td>224</td>
<td>NFTK</td>
<td>249</td>
<td>NGTD</td>
<td>304</td>
</tr>
<tr>
<td>Pig</td>
<td>Sus scrofa</td>
<td>45</td>
<td>NQSE</td>
<td>68</td>
<td>NVTN</td>
<td>105</td>
<td>NGTT</td>
<td>206</td>
<td>NGTN</td>
<td>224</td>
<td>NFTK</td>
<td>249</td>
<td>NGTD</td>
<td>304</td>
<td>NTSI</td>
<td>325</td>
</tr>
<tr>
<td>Rabbit</td>
<td>Oryctolagus cuniculus</td>
<td>45</td>
<td>NGTE</td>
<td>68</td>
<td>NVTN</td>
<td>105</td>
<td>NGTT</td>
<td>122</td>
<td>NQSV</td>
<td>206</td>
<td>NGTN</td>
<td>224</td>
<td>NFTK</td>
<td>249</td>
<td>NGTD</td>
<td>304</td>
</tr>
<tr>
<td>Elephant</td>
<td>Loxodonta africana</td>
<td>45</td>
<td>NGTE</td>
<td>68</td>
<td>NVTN</td>
<td>105</td>
<td>NGTT</td>
<td>122</td>
<td>NLSI</td>
<td>206</td>
<td>NGTN</td>
<td>224</td>
<td>NFTK</td>
<td>249</td>
<td>NGTD</td>
<td>304</td>
</tr>
<tr>
<td>Chicken</td>
<td>Gallus gallus</td>
<td>45</td>
<td>NGTE</td>
<td>68</td>
<td>NVTN</td>
<td>105</td>
<td>NGTT</td>
<td>206</td>
<td>NGTD</td>
<td>224</td>
<td>NFTK</td>
<td>249</td>
<td>NGTD</td>
<td>304</td>
<td>NTVN</td>
<td>328</td>
</tr>
<tr>
<td>Lizard</td>
<td>Anolis carolensis</td>
<td>45</td>
<td>NGTE</td>
<td>68</td>
<td>NLTN</td>
<td>105</td>
<td>NDTE</td>
<td>206</td>
<td>NGSN</td>
<td>224</td>
<td>NFTK</td>
<td>249</td>
<td>NGTD</td>
<td>304</td>
<td>NVST</td>
<td>328</td>
</tr>
<tr>
<td>Frog</td>
<td>Xenopus tropicalis</td>
<td>45</td>
<td>NESE</td>
<td>68</td>
<td>N99TF</td>
<td>105</td>
<td>NGTE</td>
<td>195</td>
<td>NTIT</td>
<td>207</td>
<td>NTID</td>
<td>250</td>
<td>NGTD</td>
<td>305</td>
<td>NVSV</td>
<td>329</td>
</tr>
<tr>
<td>Zebrafish</td>
<td>Danio rerio</td>
<td>35</td>
<td>NNTV</td>
<td>46</td>
<td>NGTE</td>
<td>106</td>
<td>NTR</td>
<td>123</td>
<td>NMSR</td>
<td>204</td>
<td>NGTE</td>
<td>247</td>
<td>NGTD</td>
<td>326</td>
<td>NASV</td>
<td></td>
</tr>
</tbody>
</table>

*Journal of Molecular Biochemistry, 2012*
in the Leu475-Ile476-Arg477 C-terminal cytoplasmic sequence (Ogata & Fukuda 1994, Tabuchi et al. 1997). The deletion of the nine amino acids closer to the SCARB2 C-terminal transmembrane domain also abolished this lysosomal location suggesting that an extended cytoplasmic tail for this protein is required for this function. Figure 1 shows a comparison of vertebrate SCARB2 C-terminal sequences, with the 7 sequences examined showing a Leu-Ile or a Leu-Leu (zebrafish SCARB2) C-terminal sequence in each case, which is consistent with a proposal for a dileucine (or leucine-isoleucine) lysosomal sorting motif reported for the COOH tails of SCARB2 (Sandoval et al. 2000) and GLUT4 (insulin-regulatable glucose transporter) (James et al. 1989). Tabuchi and coworkers (2000; 2002) have also reported that two acidic amino acids (Asp470-Glu471) in the COOH-terminal SCARB2 sequence play important roles in regulating the movement of this protein within the endocytic pathway. A comparison of vertebrate SCARB2 C-terminal sequences supports this hypothesis with acidic amino acids being predominantly conserved for these positions, with the exception of zebrafish SCARB2, which has an apparent human 470Asp→zebrafish 466Ala substitution, but with an additional acidic amino acid (469Glu) further down the C-

Figure 2. Amino Acid Sequence Alignments for Rat (Ra) SCARB2 (SCA2), SCARB1 (SCA1) and CD36 Sequences with Lancelet (La), Sea Squirt (Ss) and Fruit Fly (DmEMP1) CD-36 Like Sequences. See Table 1 for sources of rat SCARB2, SCARB1 and CD36 sequences as well as other CD36-like sequences; * shows identical residues for CD36-like 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 highlighted in green; free-SH Cys involved in lipid transfer for rat SCARB1 is highlighted in pink; predicted disulfide bond Cys residues are highlighted in cyan blue; predicted α-helices for vertebrate CD36-like sequences are highlighted in yellow and numbered in sequence from the start of the predicted exoplasmic domain; predicted β-sheets are highlighted in grey and also numbered in sequence; bold underlined font shows residues corresponding to known or predicted exon start sites; exon numbers are shown; AKL refers to final three C-terminal residues for rat SCARB1 which bind a PDZ domain-containing protein (PDZK1); transmembrane conserved glycines are shown as G; TSP1-A, TSP1-B and TSP1-C represent motifs identified by Crombie & Silverstein 1998.
terminal cytoplasmic tail (Figure 1). The intracellular trafficking of lysosomal membrane proteins has been extensively investigated in recent years, and these studies have shown that sorting from the Golgi or the plasma membrane into the endosomes and lysosomes represents one pathway which is mediated by short COOH-terminal cytoplasmic sequences (Hunziker & Geuze 1996). The dileucine cytoplasmic signal is one such pathway, however others also serve to target lysosomal proteins, such as cholesterol ester lipase (LIPA) which contains a C-terminal Arg-Lys dipeptide sequence (Sleat et al. 2006) and the C-terminal tyrosine based lysosomal targeting signal reported for lysosome-associated membrane glycoprotein-1 (LAMP-1) (Akasaki et al. 2010, Höning et al. 1996).

Predicted Secondary Structures for Vertebrate SCARB2
Predicted secondary structures for vertebrate SCARB2 sequences were examined (Figure 1), particularly for the sequences external to the lysosomal membrane (residues 28-433 for rat SCARB2) (Figure 2). α-Helix and β-sheet structures were similar in each case, with a α-helix extending beyond the N-terminal and C-terminal transmembrane regions in each case: α1 and α6. A consistent sequence of predicted secondary structure was also observed for each of the vertebrate SCARB2 sequences: N-terminal cytoplasmic sequence--N-terminal transmembrane sequence--α1--β1--α2--β2--β3--β4--β5--α3--β6--α4--α5--β7--β8--β9--β10--β11--β12--β13--δ14--α6--C-terminal trans-membrane sequence--C-terminal cytoplasmic sequence. Further description of the secondary and tertiary structures for SCARB2 must await the three dimensional structure of this protein, particularly for the external lysosomal membrane region which directly binds lipids and contributes towards the organization and maintenance of the lysosomal membranes (Gamp et al. 2003; Kuronita 2012).

Figure 3. Gene Structures and Major Transcripts for the Human, Mouse and Rat SCARB2 Genes. Derived from the AceView website http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/; 31 mature isoforms are shown with capped 5’- and 3’- ends for the predicted mRNA sequences; NM refers to the NCBI reference sequence; exons are in pink; the directions for transcription are shown as 5’→3’; black squares show predicted CpG island sites at or near the 5’untranslated regions of the genes; the symbol ⊗ shows 10 predicted microRNA binding sites observed at or near the SCARB2 3’untranslated regions; sizes of mRNA sequences are shown in kilobases (kb); predicted transcription factor binding sites (TFBS) for human SCARB2 are shown: PPARA- peroxisome proliferator-activated receptor-α; ATF-activation transcription factor; IRF7-interferon regulatory factor; HAND1-heart- and neural crest derivatives-expressed protein system; EVI1-ecotropic viral integration site; NKX 2.2-
Conserved Proline and Glycine Residues within the SCARB2 External Lysosomal Membrane Domain

Figure 1S (see supplementary data) shows the alignment of 7 vertebrate SCARB2 amino acid sequences for the external lysosomal membrane 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 SCARB2 amino acids detailed previously, others were also conserved, including 16 proline residues. Prolines play a major role in protein folding and protein-protein interactions, involving the cyclic pyrrrolidine amino acid side chain, which may introduce turns (or kinks) in the polypeptide chain as well as having destabilizing effects on α-helix and β-strand conformations (MacArthur & Thornton 1991). 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 Na+/H+ exchanger, which plays a major role in cation transport (Kreiger et al. 2005). Sequential prolines (P1 and P2: Pro57-Pro58) were conserved for 6 of 7 vertebrate SCARB2 sequences examined, which may confer a distinctive conformation in this region supporting the lipid receptor functions for this protein. For three of the vertebrate SCARB2 sequences examined, four sequential proline residues were observed (chicken, frog and zebrafish: Pro57-Pro58-Pro59-Pro60) which may contribute further to the distinctive folding conformation in this region. Regions of water exposed proteins with high levels of proline residues are often sites for protein-protein interactions (Kay et al. 2000) and these residues may significantly contribute to the binding of lipids by the external lysosomal membrane region of SCARB2. Similar results have been recently reported for vertebrate SCARB1 and CD36 exoplasmic regions, however 30 and 17 conserved proline residues were observed (Rasmussen et al. 2000); several predicted N-glycosylation sites for rat SCARB2 (11 sites) (Table 2), rat SCARB1 (10 sites) and CD36 (9 sites) (Holmes & Cox 2012), of which only one was shared between these sequences (N-glycosylation site 10) (Table 2); similar thrombospondin-1 binding sites (designated as TSP1-A, TSP1-B and TSP1-C) previously reported for SCARB2, SCARB2 and CD36 (Crombie & Silverstein 1998) and similar predicted secondary structures previously identified for SCARB1 and CD36 (Holmes & Cox 2012) (Figure 1). Sequence comparisons for the TSP1 binding sites confirmed the presence of a protein kinase C consensus phosphorylation site (Crombie & Silverstein 1998) within the TSP1-A site (GPYTYR) for rat SCARB2, CD36 and SCARB1 (but with a substitution at one site: GPYVYR). The current studies demonstrated other common features of these sites; namely the predicted β-sheet secondary structures for TSP1-A (β2) and TSP1-B (β3), the absence of predicted secondary structure for TSP1-C (located between β5 and β6) and the predicted N-glycosylation sites observed for TSP1-B. The Cys384 residue, for which the free-SH group plays a major role in SCARB1-mediated lipid transport (Yua et al. 2011), was unique to SCARB1, being replaced by other residues for the corresponding SCARB2 and CD36 proteins (Gly379 and Phe383, respectively). N-terminal transmembrane glycine residues, which play a role in the formation of SCARB1 oligomers (Gaidukov et al. 2011), were also observed for the rat CD36 sequence, although twin-glycines (Gly23-Gly24) were observed for the vertebrate CD36 sequences (Holmes & Cox 2012). In contrast, only one of these glycines (Gly10)
was observed for the rat SCARB2 N-terminal transmembrane sequence. These results suggest that rat SCARB2, SCARB1 and CD36 proteins share several important properties, features and conserved residues, including being membrane-bound with cytoplasmic and transmembrane regions, N-glycosylated at specific sites and have similar secondary structures but each is sufficiently different to serve distinct functions external to the respective membrane surfaces.

Alignments were also prepared for the predicted lancelet (Branchiostoma floridae) and sea squirt (Ciona intestinalis) CD36-like sequences (Table 1) and a major epithelial membrane protein (EMP) from the fruit fly (Drosophila melanogaster) (Nichols & Vogt 2008) with the rat SCARB2, SCARB1 and CD36 sequences (Figure 2). The lancelet, sea squirt and fruit fly sequences examined shared many features with the CD36-like rat 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 10 in Table 2) and transmembrane glycine residues, which were observed in both the N- and C-terminal sequences, although with only a single glycine residue for the human SCARB2 sequence.

**Gene Locations and Exonic Structures for Vertebrate SCARB2 Genes**

Table 1 summarizes the predicted locations for vertebrate SCARB2 genes based upon BLAT interrogations of several vertebrate genomes using the reported human and mouse SCARB2 sequences (Fujita et al. 1992, Tabuchi et al. 1997) and the predicted sequences for other vertebrate and fruit fly CD36-like genes derived from the UC Santa Cruz genome browser (Kent et al. 2003). The predicted vertebrate SCARB2 genes were transcribed on either the positive strand (e.g. chimpanzee, rhesus monkey, marmoset, rat, chicken, lizard and zebrafish genomes) or the negative strand (e.g. human, orangutan, mouse, cow, pig, opossum, chicken and frog genomes). Figure 1 summarizes the predicted exonic start sites for human, mouse, cow, chicken, lizard, frog and zebrafish SCARB2 genes with each having 12 coding exons, in identical or similar positions to those reported for the human SCARB2 gene (Kent et al. 2003).

![Human SCARB2 GNF Expression Atlas 2 Data from U133A and GNF1H Chips](image1)

![Mouse Scarb2 GNF Expression Atlas Data from GNFIN1M Chip](image2)

**Figure 4.** Comparative Tissue Expression for Human and Mouse SCARB2 Genes. Expression ‘heat maps’ (GNF Expression Atlas 2 data) (Su et al, 2004) were examined for comparative gene expression levels among human and mouse tissues for SCARB2 genes showing high (red); intermediate (black); and low (green) expression levels. Derived from human and mouse genome browsers (http://genome.ucsc.edu) (Kent et al. 2003).
Figure 3 shows the predicted structures for the major human, mouse and rat SCARB2/Scarb2 transcripts (Thierry-Mieg & Thierry-Mieg 2006). The transcripts were ~2kbs in length with 12 exons present for the mRNA transcripts and in each case, a non-coding 5’-untranslated sequence and an extended 3’-untranslated region (UTR) were observed. The human SCARB2 genome sequence contained several predicted transcription factor binding sites (TFBS), including PPARα (peroxisome proliferator-activated receptor-α), which plays a major role in kidney proximal tubule development and maintenance (Kamijo et al. 2002) where SCARB2 is highly expressed (Berkovic et al. 2008, Desmond et al. 2011); NRSF (neuron-restrictive silencing factor), which functions as a neuronal cell repressor (Kim et al. 2004); ATF (activating transcription factor 1) which mediates heme oxygenase induction by heme and drives macrophage adaptation to intraplaque hemorrhage during atherosclerosis (Boyle et al. 2012); IRF7 (interferon regulatory factor-7) which is critical for the regulation of inflammatory responses in the central nervous system (Salem et al. 2011); HAND1 (heart- and neural crest derivatives-expressed protein system) which drives ongoing expression of cardiac-specific genes (Riley et al. 1998); EVI1 (ectotropic viral integration site), which is a complex transcription factor with multiple functions (Buonamici et al. 2003) and NKX22 (homeobox protein Nkx2.2) which contributes to the expression of genes that play a role in axonal guidance (Holz et al. 2010).

Figure 3 also reports the presence of a CpG island within the promoter region of the human, mouse and rat SCARB2 genes (CpG100, CpG69 and CpG60, respectively). CpG islands were also observed within the promoter regions of other vertebrate SCARB2 genes, including marmoset (CpG106), pig (CpG71) and lizard (CpG420) (data not shown). CpG islands are typically found within the gene promoter for housekeeping genes (Saxanov et al. 2006). Elango & Yi (2011) have also proposed that larger CpG islands are associated with gene promoters showing a broad range of gene expressions and contain more RNA polymerase II binding sites than other promoters. Conse-

Figure 5: Phylogenetic Tree of Vertebrate SCARB2 Amino Acid Sequences with Human, Mouse, Chicken and Zebrafish SCARB1 and CD36 Sequences. The tree is labeled with the SCARB-like name and the name of the animal and is ‘rooted’ with the lancelet CD36 sequence. Note the 3 major clusters corresponding to the SCARB2, SCARB1 and CD36 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 replicate values of 95 or more, which are highly significant are shown with 100 bootstrap replicates performed in each case.
sequently, the presence of CpG100 and the transcription factor binding sites observed within the SCARB2 gene may contribute significantly to the broad tissue expression observed for SCARB2 transcripts. The human SCARB2 transcript also contained an extended 3’-noncoding segment with ten predicted miRNA binding sites (miR-340, 203, 10, 182, 590, 876, 141, 542, 219 and 494), which is well in excess of the usual number of such sites. miRNAs have been reported to function as post-transcriptional regulators that bind to complementary sequences on target messenger RNA transcripts (mRNAs), which result in translational repression or target degradation and gene silencing (Bartel 2009).

Comparative Human and Mouse SCARB2 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) (Su et al. 2004). These data support a broad and high level of tissue expression for human and mouse SCARB2, particularly for adipose tissue, cardiac myocytes, skeletal muscle and liver which is consistent with previous reports for these genes (Calvo et al. 1995, Fujita et al. 1992, Ogata et al. 1994, Tabuchi et al. 1997). Much lower levels of mouse Scarb2 gene expression were observed in oocytes and during early embryonic development. Overall, however, human and mouse SCARB2 tissue expression levels were 2-4 times higher than the average level of gene expression, which supports the key role played by this protein in lysosomal and endosomal membranes of the body.

Phylogeny of vertebrate SCARB2 and related CD36-like sequences

A phylogenetic tree (Figure 5) was calculated by the progressive alignment of 17 vertebrate SCARB2 amino acid sequences with human, mouse, chicken and zebrafish SCARB1 and CD36 sequences ‘rooted’ with the lancelet (Branchiostoma floridae) CD36 sequence (see Table 1). The phylogram showed clustering of the SCARB2 sequences into groups which were consistent with their evolutionary relatedness as well as groups for human, mouse, chicken and zebrafish SCARB1 and CD36 sequences, which were distinct from the lancelet CD36 sequence. These groups were significantly different from each other (with bootstrap values of ~ 100/100). It is apparent from this study of vertebrate CD-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 pre-dated the appearance of fish > 500 million years ago (Donohue & Benton 2007). In parallel with the evolution of SCARB2 and other CD36-like proteins (SCARB1 and CD36), thrombospondins (TSPs) are also undergoing evolutionary changes in their structures and functions (Bentley & Adams 2010), with gene duplication events proposed at the origin of deuterostomes.

Conclusions

The results of the present study indicate that vertebrate SCARB2 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 CD36) previously studied (Acton et al. 1996, Bultel-Briere et al. 2002, Connelly et al. 2004, Fujita et al. 1992, Holmes & Cox 2012, Kent et al. 2011, Kuronita et al. 2002, Lin et al. 2012, Marsche et al. 2003, Ogata & Fukuda 1994, Tabuchi et al. 1997). SCARB2 has a unique property among these proteins in serving major roles within endosomal and lysosomal membranes of various cells and tissues of the body. SCARB2 is encoded by a single gene among the vertebrate genomes studied and is highly expressed in human and mouse tissues, particularly in adipose tissue, cardiac myocytes, skeletal muscle and liver, and usually contained 12 coding exons. Predicted secondary structures for vertebrate CD36 proteins showed strong similarities with other CD36-like proteins, SCARB1 and CD36. Three major structural domains were apparent for vertebrate SCARB2, including the N-terminal and C-terminal cytoplasmic domains, the N-terminal and C-terminal transmembrane domains, the external lysosomal membrane domain, two disulfide bridges and several N-glycosylation sites for glycan binding, which are apparently essential for membrane recruitment. Phylogenetic studies using 17 vertebrate SCARB2 sequences with human, mouse, chicken and zebrafish SCARB1 and CD36 sequences indicated that the CD36 gene has 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.

Conflicts of Interest and Responsibility for Contents

The author declares no conflicts of interest and is fully responsible for the writing and completion of the study undertaken.
Acknowledgements

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

References


Kim JW, Zhao SH, Uthe JJ, Bearson SM & Tuggle CK 2006 Assignment of the scavenger receptor class B, member 2 gene (SCARB2) to porcine chromosome 8q11→8q12 by somatic cell and radiation hybrid panel mapping. *Cytogeten Genome Res* **112** 342H.


Papale GA, Hanson PJ & Sahoo D 2011 Extracellular disulfide bonds support scavenger receptor class B type cholesterol translocation. Biochem 50 6245–6254.


Ren Y 2012 Peroxisome proliferator-activator receptor γ: a link between macrophage CD36 and inflammation in malaria infection. PPAR Res 2012 640769.


Tabuchi N, Akasaki K & Tsuji H 2000 Two acidic amino acid residues, Asp(470) and Glu(471), contained in the carboxyl cytoplasmic tail of a major lysosomal membrane protein, LGBP5/LIMP II, are important for its accumulation in secondary lysosomes. Biochim Biophys Res Commun 257 560–565.

Tabuchi N, Akasaka K & Tsuji H 2002 LIMP II is a constituent of δ-ile-lysine-based motif of a major lysosomal membrane protein, LGBP5/LIMP II, on the bone sialoprotein, is important for its proper distribution in late endosomes and lysosomes. Biochim Biophys Res Commun 295 149–156.


Tserentsoodol N, Gordiyenko NV, Pascual I, Lee JW, Fliesser SJ & Rodriguez IR 2006 Intraretinal lipid transport is dependent on high density lipoprotein receptor class B scavenger receptors. Mol Vis 12 1319–1333.


