Mammalian carboxylesterase 5: Comparative biochemistry and genomics

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

Published
2008

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

DOI
https://doi.org/10.1016/j.cbd.2008.05.002

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

Downloaded from
http://hdl.handle.net/10072/23482
REVIEW

**Mammalian Carboxylesterase 5: Comparative Biochemistry and Genomics**

Roger S Holmes¹⁴, Laura A Cox¹² and John L VandeBerg¹²

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

⁴Corresponding Author:

Roger S Holmes, D.Sc.
Department of Genetics
Southwest National Primate Research Center
Southwest Foundation for Biomedical Research
San Antonio, TX, USA 78227

Email: rholmes@sfbrgenetics.org
Phone: 210-258-9687
Fax:210-258-9600

Keywords: Mammals; amino acid sequence; carboxylesterase

Running Head: Mammalian Carboxylesterase 5: comparative biochemistry and genomics
Summary

Carboxylesterase 5 (CES5) (also called cauxin or CES7) is one of at least five mammalian CES gene families encoding enzymes of broad substrate specificity and catalysing hydrolytic and transesterification reactions. *In silico* methods were used to predict the amino acid sequences, secondary structures and gene locations for CES5 genes and gene products. Amino acid sequence alignments of mammalian CES5 enzymes enabled identification of key CES sequences previously reported for human CES1, as well as other sequences that are specific to the CES5 gene family, which were consistent with being monomeric in subunit structure and available for secretion into body fluids. Predicted secondary structures for mammalian CES5 demonstrated significant conservation with human CES1 as well as distinctive mammalian CES5 like structures. Mammalian CES5 genes are located in tandem with the CES1 gene(s), are transcribed on the reverse strand and contained 13 exons. CES5 has been previously reported in high concentrations in the urine (cauxin) of adult male cats, and within a protein complex of mammalian male epididymal fluids. Roles for CES5 may include regulating urinary levels of male cat pheromones; catalysing lipid transfer reactions within mammalian male reproductive fluids; and protecting neural tissue from drugs and xenobiotics.

Keywords: Mammals; amino acid sequence; genomics; carboxylesterase; CES5; drug detoxification.

Running Head: Mammalian Carboxylesterase 5: comparative biochemistry and genomics
Background

Carboxylesterases (CES, E.C.3.1.1.1) catalyse hydrolytic and transesterification reactions using a broad range of substrates, detoxify organophosphates, carbamate compounds and insecticides (Ahmad & Forgash, 1976; Leinweber, 1987; Satoh and Hosokawa, 1998; Satoh et al., 2002; Redinbo and Potter, 2005), catalyse several cholesterol and fatty acid metabolic reactions (Tsujita and Okuda, 1993; Becker et al., 1994; Diczfalusy et al., 2001; Dolinsky et al., 2001), and facilitate the conversion of lung alveolar surfactant (Krishnasamy et al., 1998; Ruppert et al., 2006). CES is also involved in the biotransformation of many drugs and prodrugs (Ahmad et al., 1999; He et al., 1995; Imai et al., 2003; Ohtsuka et al., 2003; Imai, 2006; Mutch et al., 2007), anti-tumor drugs (Humerickhouse et al., 2000; Xu et al., 2002; Tabata et al., 2004) and narcotics, such as cocaine and heroine (Pindel et al., 1997), and has been linked with the assembly of very-low density lipoproteins in liver (Wang et al., 2007).

At least five families of mammalian CES have been described which show 39-45% sequence identities (Holmes et al., 2008): CES1 and CES2, the major liver and intestinal enzymes, respectively, which are both widely distributed in mammalian tissues (Shibata et al., 1993; Schewer et al., 1997; Pindel et al., 1997; Ghosh, 2000); CES3, expressed in liver, colon and brain (Sanghani et al., 1997); CES6, a predicted CES-like enzyme in brain (Clark et al., 2003); and CES5, a major urinary protein of the domestic cat (Miyazaki et al., 2003). The latter enzyme is also called cauxin (for carboxylesterase-like urinary excreted protein) or CES7 and is present in high levels in the urine of male domestic cats (Felis catus), bobcat (Lynx rufus) and lynx (Lynx lynx) (Miyazaki et al., 2003; 2006a) and also found in mammalian male reproductive fluids (Ecroyd et al., 2006). Domestic cat CES5 apparently functions in the production of felinine, a putative pheromone precursor, which has been proposed to play a role in attracting females and as a territorial marker (Miyazaki et al., 2006 b), whereas mammalian fluid CES5 may assist with lipid transport within high-molecular weight complexes in male reproductive tracts (Ecroyd et al., 2006).

This review summarizes current knowledge about mammalian CES5 and describes the predicted amino acid sequences and secondary structures and the structural, phylogenetic and evolutionary relationships for this CES gene family. The paper also reviews the major roles for mammalian CES1, CES2 and CES5 family members, and discusses potential roles for CES5 as a body fluid carboxylesterase and in contributing to a protective role from drugs and xenobiotics for the brain.

*In silico* mammalian CES5 gene and protein identification.
BLAST (Basic Local Alignment Search Tool) in silico studies were undertaken using the National Center for Biotechnology Information (NCBI) web site [http://www.ncbi.nlm.nih.gov/blast/Blast.cgi] for nucleotide and protein blasts using known CES sequences (Table 1) (Altschul et al., 1990). BLAT (BLAST-Like Alignment Tool) in silico studies were also undertaken using the UC Santa Cruz Genome Browser [http://genome.ucsc.edu/cgi-bin/hgBlat] (Kent et al. 2002) with the default settings. Genomes examined included: human (Homo sapiens) (International Genome Sequencing Consortium, 2004; chimpanzee (Pan troglodytes); (The Chimpanzee Sequencing and Analysis Consortium, 2005); orangutan (Pongo abelii) (Orangutan Genome Sequencing project, 2008); rhesus monkey (Macaca mulatta) (Rhesus Macaque Genome Sequencing and Analysis Consortium, 2007); mouse (Mus musculus) (Mouse Sequencing Consortium, 2002); rat (Rattus norvegicus) (Rat genome Sequencing project Consortium, 2004); cat (Felis catus) (Pontius et al., 2007); dog (Canis familiaris) (Dog Genome Project., 2005); cow (Bos taurus) (Bovine Genome Project, 2008); and horse (Equus caballus) (Horse Genome project, 2008).

UniProtKB/Swiss-Prot Database [http://au.expasy.org] and GenBank [http://www.ncbi.nlm.nih.gov/Genbank/] sequences for human, mouse (Miyazaki et al., 2006a), rat, cat (Miyazaki et al., 2003), dog and sheep (Ecroyd et al., 2006) CES5 were used to interrogate known mammalian genome sequences (Table 1). Gene locations, predicted gene structures and CES protein subunits were observed for each CES sequence examined for those regions showing identity with the respective mammalian CES5 gene products.

**Predicted Secondary Structures for Mammalian CES5**

Predicted secondary structures for human CES2 and various mammalian CES5 subunits were obtained using the PSIPRED v2.5 web site tools [http://bioinf.cs.ucl.ac.uk/psipred/psiform.html] (McGuffin et al., 2000).

**Phylogenetic Studies and Sequence Divergence**

Phylogenetic trees were constructed using an amino acid alignment from a ClustalW-derived alignment of CES protein sequences, obtained with default settings and corrected for multiple substitutions (Chenna et al 2003) [http://www.ebi.ac.uk/clustalw/]. Alignment ambiguous regions, including the amino and carboxyl termini, were excluded prior to phylogenetic analysis yielding an alignment of 526 residues. Amino acid sequence divergences for mammalian CES5 were determined using the SIM-Alignment tool for Protein Sequences [http://au.expasy.org/tools/sim-prot.html] (Pietsch 1995; Schwede et al. 2003).

**Alignments of human CES1, CES2 and CES5 amino acid sequences.**

The amino acid sequences for human CES1 (Shibata et al., 1993; Becker et al., 1994; Ghosh, 2000) and CES2 (Schewer et al., 1997; Pindel et al., 1997) and the predicted sequence for human CES5 (International Human
Genome Sequencing Consortium, 2004) are shown in Figure 1 (see Table 1). Alignments of human CES1 with human CES2 and CES5 showed 45% and 42% sequence identities respectively, while the alignment of human CES2 with CES5 showed 43% sequence identity (Table 2). This comparison of human CES1, CES2 and CES5 subunit sequences suggests that they are products of separate families of CES genes, which supports earlier proposals (Satoh and Hosokawa, 2006; Holmes et al., 2008).

The predicted amino acid sequence for human CES5 was eight residues longer (575 amino acids) than for human CES1 (567), and 16 residues longer than for human CES2 (559 amino acids) (Figure 1). Comparisons of key residues are of particular interest, which have been previously shown to contribute to the catalytic, subcellular localization, oligomeric and regulatory functions for human CES1 (sequence numbers refer to human CES1). These include the active site catalytic triad (Ser228; Glu345; His458) (Cygler et al., 1993); microsomal targeting sequences, including the hydrophobic N-terminus signal peptide (von Heijne 1983; Zhen et al., 1995; Potter et al., 1998) and the C-terminal endoplasmic reticulum (ER) retention sequence (His-Ile-Glu-Leu) (Robbi & Beaufay 1983; Munro & Pelham 1987; Zhen et al., 1995); disulfide bond forming residues (Cys95/Cys123 and Cys280/Cys291) (Lockridge et al., 1987); and ligand binding sites, including the ‘Z-site’ (Gly358), the ‘side door’ (Val424-Met425-Phe426) and the ‘gate’ (Phe551) residues (Redinbo and Potter, 2005). Other CES class 1 specific residues included the ‘charge clamp’ residues contributing to the oligomeric subunit structures for human CES1 (Lys78..Glu183 and Glu72..Arg193), and the N-glycosylation site at Asn79-Ala80-Thr81 (Ozols, 1989; Kroetz et al, 1993; Bencharit et al., 2003; 2006; Fleming et al., 2005). The predicted human CES5 sequence contained essential catalytic and structural residues, such as the active site triad, the disulfide bridges and the ‘Z-site’, and exhibited conservative changes in sequences for other key residues, such as the ‘side door’ and ‘gate’, which are proposed to facilitate product (acyl group) release from the CES active site (Redinbo and Potter, 2005).

Human CES5 showed a distinct predicted N-terminal ‘signal peptide’ sequence to that reported for human CES1 and CES2, and lacked relevant charge clamp residues, which contribute to the human CES1 oligomeric structure (trimer-hexamer) (Fleming et al., 2005). Human and baboon CES2 have been shown to be monomeric enzymes (Pindel et al., 1997; RS Holmes, J Glenn, JL Vandeberg and LA Cox, Baboon carboxylesterases 1 and 2: sequences , structures and phylogenetic relationships with human and other primate carboxylesterases] and it is likely that human CES5 is also monomeric due to the absence of charge clamps reported for human CES1 (Fleming et al, 2005). The predicted C-terminus for human CES5 lacked the reported endoplasmic reticulum retention sequence human CES1 and CES2 (HIEL and HTEL, respectively), which may contribute to this enzyme being secreted, rather than being localized within the endoplasmic reticulum (Clark et al., 2003). Four potential N-
glycosylation sites were observed for human CES5 (Table 3) which may assist in stabilizing the enzyme in support of its location in body fluids such as urine (observed in cats) (Miyazaki et al., 2006a) and in male reproductive fluids (observed in sheep) (Ecroyd et al., 2005). N-glycosylation at the single Asn79-Ala80-Thr81 human CES1 site has been previously reported to contribute to the stability of the enzyme and in maintaining maximal catalytic for this enzyme (Kroetz et al., 1993; Fleming et al., 2005), and similar roles for the multiple N-glycosylation human CES5 sites are proposed.

**Comparative Mammalian CES Genomics**

Structures for several mammalian CES genes have been determined, including human (Becker et al., 1994; Langmann et al., 1997; Pindel et al., 1997; Ghosh 2000; Marsh et al., 2004) and rodent CES1 and CES2 ‘like’ genes (Ghosh et al., 1995; Dolinsky et al., 2001; Hosokawa et al., 2007; Furihata et al., 2003; 2005; 2006).

Six human CES genes have been described on chromosome 16 (Figure 2). The human CES1 gene, encoding the major liver CES but also expressed in lung, tumours, oesophagus, breast, uterus and most other tissues of the body (Shibata et al., 1993; Becker et al., 1994), is defined by > 600 GenBank entries (Thierry-Mieg and Thierry-Mieg, 2006) and maps at 16q13-q22.1 on the reverse strand (Table 1). The CES1 gene is transcribed into 5 alternatively spliced mRNAs, of which 3 are derived from 14 exons, encoding CES protein subunits that appear to be fully functional, forming CES1a, CES1b and CES1c isoforms (Figure 3). The human CES2 gene encodes the major intestinal CES (Schewer et al., 1997; Pindel et al., 1997) and is also expressed in most other tissues of the body, including liver, colon, kidney, brain and lung (Thierry-Mieg and Thierry-Mieg, 2006). The CES2 gene sequence is defined by > 500 GenBank entries and maps at 16q22.1 on the positive strand (Table 1). CES2 transcription produces 11 alternatively spliced mRNAs, of which at least 3 are derived from 12 exons and form apparent fully functional CES subunits, namely CES2a, CES2b and CES2c (Figure 2).

The human CES5 gene sequence is currently defined by 17 GenBank entries and encodes at least 4 alternatively spliced mRNAs which are derived from several tissues of the body, including brain, testis and lung (International Human Genome Sequencing Consortium, 2004; Thierry-Mieg and Thierry-Mieg, 2006) (Figure 3). Human CES5 maps on chromosome 16 at 16q12.2 in a tandem location to human CES1 and encodes 4 isoforms which appear to be fully functional, derived from 13 exons: CES5a, CES5b, CES5c and CES5d. Figure 2 describes the gene structures for human CES1, CES2 and CES5, based on the 5’ to 3’ alignments of multiple mRNAs respectively on the human genome and obtained using AceView (Thierry-Mieg and Thierry-Mieg, 2006). These genes are transcribed into several multiple splice variants in each case, although only those isoforms with capped 5’
ends and validated 3’ ends are shown: CES1 encodes at least 3 such isoforms with 14 exons in each case; human CES2 encodes at least 3 mature isoforms although these are derived from 12 exons; while human CES5 encodes at least 4 mature isoforms with the CES5a isoform being derived from 13 exons. Human CES1 and CES5 are encoded on the minus strand whereas human CES2 is encoded on the plus strand.

Figure 1 shows the locations of the intron-exon boundaries for human CES1, CES2 and CES5 gene products and their positioning within the aligned amino acid sequences. Exon 1 corresponds to the encoded signal peptide in each case, with the last exon encoding the endoplasmic reticulum targeting sequence (for human CES1 and CES2) or the C-terminal sequence involved in a proposed role in the secretion of human CES5. There is identity or near identity for the intron-exon boundaries for each of the human CES1, CES2 and CES5 genes, with the exception of an additional human CES1 exon boundary (forming exon 9), and the absence of an intron-exon boundary for human CES2 within exon 5.

At least 3 additional human CES genes have been described which are also located on chromosome 16 within two segments of CES gene clusters (Figure 2; Table 1). The human CES3 gene sequence is currently defined by 65 GenBank entries which span 13.93 kb on chromosome 16 downstream from CES2 (Sanghani et al., 2004; Thierry-Mieg and Thierry-Mieg, 2006); the human CES4 gene (Yan et al., 1999) is complex in nature, being transcribed into 4 alternatively spliced mRNAs, which may form multiple CES proteins with no sequence overlap and located near the CES1 gene on chromosome 16 (Thierry-Mieg and Thierry-Mieg, 2006); whereas the human CES6 gene (also called FLJ37464) sequence (Clark et al., 2003) is defined by 86 GenBank entries which are derived from mRNAs isolated from >40 tissues of the body and is located downstream of CES2 and CES3 on chromosome 16 (Table 1; Figure 2) (Thierry-Mieg and Thierry-Mieg, 2006).

**Secondary Structures for Human CES1, CES2 and CES5**

Figure 1 shows the comparative secondary structures previously reported for human CES1 (Bencharit et al., 2003; 2006; Fleming et al, 2005) or predicted for human CES2 and CES5. Similar α-helix β-sheet structures were observed for the three human CES gene products examined, which was readily apparent near key residues or functional domains, including the α-helix within the N-terminal signal peptide and the β-sheet and α-helix structures surrounding the active site Ser228 (human CES1). In addition, a large random coil region (residues 55-117 for human CES1) was predominantly retained within human CES2 and CES5. Other sites however showed differences in predicted secondary structures for human CES5, in comparison with the reported structure for human CES1, including the predicted ‘Z-site’ for CES5 (human CES1 Gly356), which has a larger helix involving both this site
and the proximate active site residue (human CES1 Glu354); the predicted ‘side door’ for CES5 (human CES1 Val424-Met425-Phe426), for which a large helix (human CES1 residues 410-439) was divided into 2 smaller predicted helices for human CES5; and an additional $\alpha$-helix which was observed at the extended C-terminus for human CES5. It should be recognized however that predictions of CES secondary structure may not fully reflect CES structures in vivo and may serve only as a guide as to the comparative structures for CES5 and other CES subunits.

Comparative Amino Acid Sequences for Mammalian CES5. A Stop Codon within the Cat CES5 Gene.

Figure 4 examines the comparative amino acid sequences for CES5 from 11 mammalian species. These CES5 sequences were complete with the exception of sheep epididymus CES5 (Ecroyd et al., 2005) which aligned with residues 118-498 of human CES5; and cat CES5, reported by Miyazaki et al (2003) and based on a cat kidney RT-PCR cDNA clone, which had a sequence aligning with human CES5 residues 1-544. The cat CES5 sequence, which lacked 30 residues at the C-terminus end, has apparently arisen from a gene mutation which introduced a stop codon (TGA) at a position encoding residue 545 (normally Ile545) for other mammalian CES5 sequences. A comparison of dog and cat CES5 exon 13 DNA sequences (Figure 5) indicated that a GT insertion at this position in the cat genome created a stop codon that resulted in a shortened (544 residues) cat CES5. The reduction in length for the sheep CES5 sequence may have arisen from either posttranslational modification of the male reproductive secreted enzyme; from the translation of a shorter spliced isoform CES5 mRNA isoform; or from gene sequence changes for the sheep CES5 gene.

The following mammalian CES5 residues were consistent with functions and aligned positions to those described for human CES1 (Redinbo and Potter, 2005; Bencharit et al., 2003; 2006): a 26 residue N-terminus hydrophobic signal peptide containing a large $\alpha$-helix; a ten residue (amino acids 56-65) trans-membrane sequence; four cysteine residues forming disulfide bonds (Cys94; Cys102; Cys278; Cys291) ; CES5 active site triad residues : 226Ser; 345Glu; and 437His; a predicted ‘Z-site’ for CES7 at 347Gly, which may play a role in cholesterol binding; the predicted CES5 ‘side door’ at residues 410Val-411Phe-412Phe; and a predicted human CES5 active site ‘gate’ at 543Ile, which may play roles in facilitating the release of acyl groups following hydrolysis or transesterification.

Mammalian CES5 sequences differed in the number and distribution of potential N-glycosylation sites which may play a role in enhancing the stabilities and catalytic activities for these enzymes (Table 3). Seven such sites were observed although only three of these were found in most CES5 sequences examined, including Site3: 281Asn-282Ala/Val/Ser-282Ser; Site 4: 363Asn-364Lys/Glu-365Ser; and Site 7: 522Asn-523Met/Ile/Val-524Ser.
Most mammalian CES5 sequences contained 3 or 4 such sites which compares with the single N-glycosylation site reported for human and baboon CES1 and CES2 (Kroetz et al., Pindel et al, 1997; RS Holmes, J Glenn, JL VandeBerg and LA Cox, Baboon carboxylesterases 1 and 2: sequences, structures, and phylogenetic relationships with human and other primate carboxylesterases, unpublished). This additional capacity for N-glycosylation and binding of carbohydrate residues may contribute significantly to increasing the stability and activity of these enzymes, particularly in light of their distribution in fluids of the body.

Predicted secondary structures for the mammalian CES5 sequences examined showed similar patterns of α-helices and β-strands, although some species differences were observed, including an additional helix (residues 242-251) for orangutan and horse CES5; the absence of predicted helices following the 345Glu active site residue for mouse and rat CES5 (residues 348-354) and rat, human and chimp CES5 (residues 356-359), respectively; and the lack of a predicted C-terminal helix for rat CES5.

Phylogeny of Human CES Gene Products and of Mammalian CES5

Phylogenetic and sequence alignments studies on vertebrate and invertebrate CES have reported evidence for at least five distinct CES gene family clusters, namely CES1, CES2, CES3, CES5 and CES6, arising from rapid gene duplication events between 328-378 million years ago (Holmes et al., 2008). In addition, further gene duplication events have apparently occurred during mammalian evolution among common ancestors for rodents, generating at least four CES1 and CES2 like genes in each case; and the opossum, where three CES2 like genes have been described (RS Holmes, J Glenn, JL VandeBerg & LA Cox: Baboon carboxylesterases 1 and 2: sequences, structures and phylogenetic relationships with human and other primate carboxylesterases, unpublished; Holmes et al., 2008).

A phylogenetic tree (Figure 6) was calculated by the progressive alignment of five human CES amino acid sequences and eleven mammalian CES5 sequences which shows clustering into five main groups (or families) for the CES1, CES2, CES3, CES5 and CES6 ‘like’ gene products. The eleven mammalian CES5 sequences clustered together within a single group which supports the proposal that these form part of a single enzyme class. Table 2 summarizes the percentages of identity for these enzymes and shows that mammalian CES5 sequences are > 66% identical which contrasts with the 42-45% identities observed for human CES1, CES2 and CES5. In addition, more closely related species show higher levels of sequence identity for CES5, including > 95% identical for primate species, human, chimpanzee, orangutan and rhesus monkey. Holmes et al. (2008) have recently reported evidence for a rapid early diversification into at least five CES gene family clusters, namely CES1; CES2; CES3; CES5; and
CES6, and reported an estimated time for the origin of these CES gene duplications at around 328-378 million years ago. Based on this report, we have concluded that the CES7 primordial gene predates the eutherian mammalian common ancestor (estimated at around 105 million years ago) (Murphy et al., 2001; Woodburne et al., 2003) by >200 million years. In addition, it is also apparent that the mammalian CES5 gene has been derived from a common ancestor shared by all eutherian mammals ~100 MY ago.

**CES5 Functional Aspects**

Mammalian carboxylesterase 5 (CES5) (also called cauxin and CES7) is a distinct CES gene family and a member of the CES super-family of genes encoding enzymes of broad substrate specificity and responsible for the detoxification and metabolism of a wide range of carboxylesters, thioesters and aromatic amides, including xenobiotics, narcotics and clinical drugs, and which have the capacity to catalyze a number of cholesterol and lipid metabolic reactions (Becker et al., 1994; Satoh and Mizakawa, 1998; Satoh et al., 2002; Redinbo and Potter, 2005; Tsujita and Okuda, 1993; Diczfalusy et al., 2001; Dolinsky et al, 2001). More specific roles for mammalian CES have been described, including the activation of lung surfactant (Krishnasamy et al., 1998); the detoxification of organophosphate and carbamate poisons (Jakovivc et al., 1996; Satoh and Hosokawa, 1998); the activation of several prodrugs used in treating various diseases such as influenza (He et al., 1999), cancer (Humerickhouse et al., 2000; Ohtsuka et al., 2003; Tabata et al., 2004) and high blood pressure (Takai et al., 1997); providing additional metabolic capacity as lipid hydrolases for neutral fats and phospholipids in tissues of the body (Mentlein et al., 1988; Ghosh et al., 1995; Tsujita and Okuda, 1993; Okazaki et al., 2006); and contributing to glucocorticosteroid drug activation within bronchial cells following prodrug inhalation (Mutch et al., 2007).

The differential tissue distribution and microlocalization of CES family members may provide an important clue as to their roles within mammalian organisms. Mammalian liver is the predominant site for drug metabolism in the body and is also the major source of CES1 and CES2 (with CES1 > CES2) and is also the predominant site of drug metabolism in the body, where these enzymes play major roles in drug clearance from the body, following absorption of drugs and xenobiotics into the circulation (Pindel et al., 1997; Imai, 2006). CES1 and CES2, which are found with high activities in mammalian intestine (with CES2 > CES1), have been implicated in the first pass clearance of several drugs in rats. The predominant intestinal drug metabolism activity is located in the ileum and jejunum and processed via CES2 (Imai et al., 2006). Differential roles for CES1 and CES2 in drug metabolism have been investigated for the anti-cancer drug irinotecan (CPT-11) which is converted to its active form SN-38 predominantly by CES2 which has a 12 to 26 fold higher affinity for CPT-11 as compared with CES1 (Humerickhouse et al., 2000; Xu et al., 2002)
In contrast with mammalian CES1 and CES2, which are predominantly localized within the liver and intestine endoplasmic reticulum and are strongly membrane bound, mammalian CES5 is predominantly expressed in peripheral tissues, including brain, kidney, lung and testis (Thierry-Mieg and Thierry-Mieg, 2006), and is a secreted form of CES enzyme due to the absence of the microsomal targeting sequence found at the carboxy-terminus (Figure 3; Figure 5). CES1 HIEL and CES2 HTEL C-terminal sequences, and similar tetrapeptide sequences, have been previously shown to direct the microlocalization of mammalian liver CES within the endoplasmic reticulum (Robbi and Beaufay, 1991). Miller and coworkers (1999) have modified this sequence for human CES1 using site-directed mutagenesis to that of HIER, substituting the C-terminal leucine with arginine, and observed that the enzyme was secreted rather than retained within the cultured cells examined. Human and other mammalian CES5 sequences normally have a long hydrophobic C-terminal sequence which apparently provides the trigger for secretion. The cat CES5 amino acid sequence lacked the elongated hydrophobic C-terminal sequence observed for all other mammalian CES5 sequences examined and the C-terminus tetrapeptide motif observed for human CES1 and CES2 sequences.

Recent studies on mammalian CES5 have supported at least two major roles for this secreted enzyme within mammalian fluids. For example, domestic cat CES5 (also called CES7 and cauxin for carboxylesterase-like urinary excreted protein) represents a major protein in urine, particularly from adult males, being secreted from epithelial cells of kidney distal tubules where it is apparently functions in regulating the production of a pheromone precursor (Miyazaki et al. 2003; 2006). CES5 has also been identified in mammalian male reproductive fluids, specifically in the epididymal fluids of sheep, pigs, mice, horses and cats, where it was associated with a soluble form of the prion protein (Ecroyd et al., 2005). The physiological role for epididymyl CES5 in mammals remains to be determined however it is likely that the enzyme performs several roles given its broad substrate specificity, and may contribute significantly to lipid and cholesterol transfer processes within male reproductive fluids. With the exception of cat, mammalian CES5 sequences retained the Z-site glycine (347Gly for human CES7). This residue facilitates cholesterol analogue binding by human CES1 (Redinbo and Potter, 2005; Bencharit et al., 2003; 2006), and its retention may indicate that this property is shared by mammalian CES5 which may play a role in cholesterol ester metabolism. In addition, human CES5 ‘side door’ (410Val-411Phe-412Phe) and ‘gate’ (543Ile) residues have retained the hydrophobic property observed for the corresponding human CES1 residues, and this may reflect similar roles for mammalian CES5 in regulating acyl release following hydrolysis or transesterification (Redinbo and Potter, 2005; Bencharit et al., 2003; 2006).
CES5 has also been identified in human brain (Ota et al., 2004) and is likely to be localized in the cerebrospinal fluid or other fluids of the brain. This expression and likely distribution in neural fluids may provide a guide as to its role, such as protecting the brain and other neural tissues from drugs via the blood brain barrier or the cerebrospinal fluid. We conclude that CES5 plays a distinctly different role to that of the major liver (CES1) and intestine (CES2) enzymes, and may serve as a soluble fluid form of CES within peripheral tissues of the body, with specialized functions in drug, xenobiotic and lipid metabolism, which remain to be fully determined.

Conclusions

The results of the present study have demonstrated that mammalian CES5 from several species share key conserved sequences and structures that have been reported for human CES1, which are consistent with this family of enzymes also serving a broad role in carboxyl ester hydrolysis and transesterification reactions. Mammalian CES5 also exhibited family specific sequences which are distinct from the CES1 and CES2 gene families previously investigated (Bencharit et al., 2003; 2006; Schewer et al., 1997; Pindel et al., 1997), suggesting specialized roles in metabolism. Moreover, all of the mammalian CES5 sequences studied lacked the C-terminus sequence which ensures retention of the enzyme within the endoplasmic reticulum (HIEL and HVEL for human CES1 and CES2 respectively) enabling secretion of this enzyme into fluids of the body where it may perform specific roles in drug, pheromone and lipid metabolism.

Acknowledgements:

This project was supported by NIH Grants P01 HL028972 and P51 RR013986. In addition, this investigation was conducted in facilities constructed with support from Research Facilities Improvement Program Grant Numbers 1 C06 RR13556, 1 C06 RR15456, 1 C06 RR017515.

REFERENCES


CES1 Gene Card, 2008, GC16M054395 http://www.genecards.org/cgi-bin/carddisp.pl?gene=CES1


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


Robbi M., Beaufay H., 1991. The COOH terminus of several liver carboxylesterases targets these enzymes to the lumen of the endoplasmic reticulum. J. Biol. Chem. 266, 20498-20503.


Xu, G., Zhang, W., Ma, M.K., McLeod, H.L., 2002. Human carboxylesterase 2 is commonly expressed in tumor tissue and is correlated with the activation of irinotecan. Clin. Cancer Res. 8, 2605-2611.


Table 1: Mammalian CES7 Genes and Human CES Genes.

<table>
<thead>
<tr>
<th>Mammal</th>
<th>CES Gene</th>
<th>GenBank ID</th>
<th>UNIPROT ID</th>
<th>No of Amino Acids</th>
<th>Chromosome Location</th>
<th>Strand</th>
<th>CES Gene Order</th>
<th>Gene Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>CES5</td>
<td>BC069501</td>
<td>Q6YTI2</td>
<td>575</td>
<td>16:54,437,867-54,466,634</td>
<td>Negative</td>
<td>CES4-CES1-CES5</td>
<td>28767</td>
</tr>
<tr>
<td>Chimp</td>
<td>CES5</td>
<td>*chr16.56.003</td>
<td>575</td>
<td>16:43,074,925-43,104,130</td>
<td>Negative</td>
<td>CES1-CES5</td>
<td>31205</td>
<td></td>
</tr>
<tr>
<td>Orangutan</td>
<td>CES5</td>
<td>*chr16.470.1</td>
<td>575</td>
<td>16:19,11,915,821-11,938,412</td>
<td>Positive</td>
<td>CES5-CES4-CES1-CES3</td>
<td>27581</td>
<td></td>
</tr>
<tr>
<td>Mouse</td>
<td>CES5</td>
<td>AB186393</td>
<td>Q6AW46</td>
<td>575</td>
<td>8:96,038,096-96,059,607</td>
<td>Negative</td>
<td>CES1-CES5</td>
<td>34074</td>
</tr>
<tr>
<td>Rat</td>
<td>CES5</td>
<td>AF479659</td>
<td>Q5GRG2</td>
<td>575</td>
<td>20:54,155,029-54,189,003</td>
<td>Negative</td>
<td>CES1-CES5</td>
<td>21511</td>
</tr>
<tr>
<td>Cat</td>
<td>CES5</td>
<td>AB045377</td>
<td>Q81O34</td>
<td>575</td>
<td>18:24,405,935-24,433,935</td>
<td>Negative</td>
<td>CES1A-CES5-CES1B</td>
<td>28000</td>
</tr>
<tr>
<td>Sheep</td>
<td>CES5</td>
<td>AM075621</td>
<td>Q3T930</td>
<td>575</td>
<td>3:8,287,880-8,318,561</td>
<td>Negative</td>
<td>CES1A-1C-1D-1E-CES5</td>
<td>30681</td>
</tr>
<tr>
<td>Human</td>
<td>CES1</td>
<td>L07765</td>
<td>P23141</td>
<td>575</td>
<td>16:65,925,828-65,952,268</td>
<td>Negative</td>
<td>CES4-CES5-CES5</td>
<td>30311</td>
</tr>
<tr>
<td></td>
<td>CES2</td>
<td>BX538086</td>
<td>O00248</td>
<td>575</td>
<td>16:65,552,639-54,364,924</td>
<td>Negative</td>
<td>CES4-CES1-CES5</td>
<td>12912</td>
</tr>
<tr>
<td></td>
<td>CES3</td>
<td>NM024922</td>
<td>Q6N9W8</td>
<td>575</td>
<td>16:65,580,134-65,602,403</td>
<td>Positive</td>
<td>CES2-CES3-CES6</td>
<td>10665</td>
</tr>
<tr>
<td></td>
<td>CES4</td>
<td>BC131699</td>
<td>Q8T209</td>
<td>575</td>
<td>16:65,915,921-915,921</td>
<td>Negative</td>
<td>CES4-CES1-CES5</td>
<td>13913</td>
</tr>
<tr>
<td></td>
<td>CES6</td>
<td>FLJ37464</td>
<td>Q6UX55</td>
<td>575</td>
<td>16:65,925,828-925,828</td>
<td>Positive</td>
<td>CES2-CES3-CES6</td>
<td>22269</td>
</tr>
</tbody>
</table>
Table 2: Predicted N-Glycosylation Sites for Mammalian CES5

<table>
<thead>
<tr>
<th>Mammal</th>
<th>CES Gene</th>
<th>Site 1</th>
<th>Site 2</th>
<th>Site 3</th>
<th>Site 4</th>
<th>Site 5</th>
<th>Site 6</th>
<th>Site 7</th>
<th>No of Sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>Orangutan</td>
<td>CES5</td>
<td>281-3</td>
<td>363-5</td>
<td></td>
<td></td>
<td>511-3</td>
<td>522-4</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NAS</td>
<td>NKS</td>
<td></td>
<td></td>
<td>NLT</td>
<td>NMS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human</td>
<td>CES5</td>
<td>281-3</td>
<td>363-5</td>
<td></td>
<td></td>
<td>511-3</td>
<td>522-4</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NAS</td>
<td>NES</td>
<td></td>
<td></td>
<td>NLT</td>
<td>NMS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chimp</td>
<td>CES5</td>
<td>281-3</td>
<td>363-5</td>
<td></td>
<td></td>
<td>511-3</td>
<td>522-4</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NAS</td>
<td>NKS</td>
<td></td>
<td></td>
<td>NLT</td>
<td>NMS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rhesus</td>
<td>CES5</td>
<td>281-3</td>
<td>363-5</td>
<td></td>
<td></td>
<td>511-3</td>
<td>522-4</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NAS</td>
<td>NKS</td>
<td></td>
<td></td>
<td>NLT</td>
<td>NMS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mouse</td>
<td>CES5</td>
<td>134-6</td>
<td></td>
<td>363-5</td>
<td></td>
<td></td>
<td>522-4</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NGS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>NIS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rat</td>
<td>CES5</td>
<td>281-3</td>
<td>363-5</td>
<td></td>
<td></td>
<td></td>
<td>522-4</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NVS</td>
<td>NES</td>
<td></td>
<td></td>
<td></td>
<td>NIS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Horse</td>
<td>CES5</td>
<td>86-88</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>522-4</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>NIS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cow</td>
<td>CES5</td>
<td>86-88</td>
<td>134-6</td>
<td>363-5</td>
<td>443-5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>NAT</td>
<td>NGS</td>
<td></td>
<td></td>
<td></td>
<td>NIS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cat</td>
<td>CES5</td>
<td>82-82</td>
<td>281-3</td>
<td>363-5</td>
<td></td>
<td></td>
<td>522-4</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NAT</td>
<td>NGS</td>
<td></td>
<td></td>
<td>NKS</td>
<td>NDT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dog</td>
<td>CES5</td>
<td>263-5</td>
<td>235-7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NDS</td>
<td></td>
<td></td>
<td></td>
<td>NKS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sheep</td>
<td>CES5</td>
<td>263-5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NAS</td>
<td>NKS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Numbers refer to amino acids in CES5 Sequence: N-Asn; A-Ala; S-Ser; V-Val; T-Thr; G-Gly; K-Lys; L-Leu; E-Glu; D-Asp; I-Ile; M-Met.
Table 3: Percentage Identities for Mammalian CES5 Amino Acid Sequences

<table>
<thead>
<tr>
<th>CES</th>
<th>hCES5</th>
<th>chCES5</th>
<th>orCES5</th>
<th>rhCES5</th>
<th>mCES5</th>
<th>raCES5</th>
<th>caCES5</th>
<th>doCES5</th>
<th>coCES5</th>
<th>hoCES5</th>
<th>shCES5</th>
</tr>
</thead>
<tbody>
<tr>
<td>hCES5</td>
<td>100</td>
<td>98</td>
<td>96</td>
<td>95</td>
<td>69</td>
<td>69</td>
<td>70</td>
<td>75</td>
<td>74</td>
<td>76</td>
<td>80</td>
</tr>
<tr>
<td>chCES5</td>
<td>98</td>
<td>100</td>
<td>96</td>
<td>96</td>
<td>70</td>
<td>70</td>
<td>70</td>
<td>75</td>
<td>74</td>
<td>76</td>
<td>80</td>
</tr>
<tr>
<td>orCES5</td>
<td>96</td>
<td>96</td>
<td>100</td>
<td>94</td>
<td>68</td>
<td>69</td>
<td>70</td>
<td>86</td>
<td>64</td>
<td>68</td>
<td>67</td>
</tr>
<tr>
<td>rhCES5</td>
<td>95</td>
<td>96</td>
<td>94</td>
<td>100</td>
<td>70</td>
<td>70</td>
<td>70</td>
<td>75</td>
<td>69</td>
<td>75</td>
<td>80</td>
</tr>
<tr>
<td>mCES5</td>
<td>69</td>
<td>70</td>
<td>68</td>
<td>70</td>
<td>100</td>
<td>86</td>
<td>64</td>
<td>68</td>
<td>67</td>
<td>66</td>
<td></td>
</tr>
<tr>
<td>raCES5</td>
<td>69</td>
<td>70</td>
<td>69</td>
<td>70</td>
<td>67</td>
<td>100</td>
<td>83</td>
<td>74</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>caCES5</td>
<td>70</td>
<td>70</td>
<td>69</td>
<td>69</td>
<td>64</td>
<td>65</td>
<td>100</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>doCES5</td>
<td>75</td>
<td>75</td>
<td>74</td>
<td>75</td>
<td>68</td>
<td>69</td>
<td>83</td>
<td>83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>coCES5</td>
<td>74</td>
<td>74</td>
<td>73</td>
<td>74</td>
<td>67</td>
<td>67</td>
<td>74</td>
<td>81</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hoCES5</td>
<td>76</td>
<td>76</td>
<td>75</td>
<td>75</td>
<td>66</td>
<td>67</td>
<td>78</td>
<td>85</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>shCES5</td>
<td>80</td>
<td>80</td>
<td>78</td>
<td>80</td>
<td>73</td>
<td>72</td>
<td>75</td>
<td>83</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

h-human; ch-chimpanzee; or-Orangutan; rh-Rhesus monkey; m-mouse; ra-rat; ca-cat; do-dog; co-cow; ho-horse; sh-sheep.

Figure 1: Amino acid sequence alignments for human CES1, CES2 and CES5.

See Table 1 for sources of CES sequences; * shows identical residues for human CES1, CES2 and CES5; : 2 alternate residues. Residues involved in endoplasmic reticulum processing at N-(Signal peptide) and C- termini (MTS-microsomal (endoplasmic reticulum) targeting sequence); N-glycosylation residues at 79NAT (Human CES1) and potential N-glycosylation sites; shows active site triad residues Ser; Glu; and His. 'Side door', 'Gate' residues and Cholesterol binding Gly residue (Z site) for human CES1 Disulfide bond ---- Cys residues for human CES1. Charge clamp residues identified for human CES1; Helix (Human CES1 or predicted helix; Sheet (Human CES1) or predicted sheet. Large font shows known or predicted exon junctions. Exons are numbered for CES1, CES2 and CES5. [Predicted transmembrane regions are shown]
Figure 2: Locations of CES Genes on Human Chromosome 16. Numbers refer to kilobases of DNA.

Figure 3: Gene Structures and Slicing Variants for Human CES1, CES2 and CES5 Genes. Derived from the AceView website (Thierry-Mieg and Thierry-Mieg, 2006). http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/av.cgi?exdb=AceView&db=36&term=CES7&submit=G

Mature isoform variants (a, b, c etc) are shown with capped 5'- and validated 3'-ends for the predicted mRNA sequences. NM numbers refer to annotated RefSeq sequences for CES1, CES2 and CES5 genes. Scale refers to base pairs of nucleotide sequences.

Figure 4: Amino acid sequence alignments for mammalian CES5.

See Table 1 for sources of CES5 sequences; * show identical residues for mammalian CES5; : 2 alternate residues. Residues involved in microsomal processing at N- (Signal peptide); N-glycosylation residues at potential N-glycosylation sites; □ shows active site triad residues □□□ and □□□. 'Side door', 'Gate' residues and Cholesterol binding Gly residue (Z site) (for human CES1) Disulfide bond ---- residues. Charge clamp residues previously identified for human CES1; Predicted helix; Predicted Sheet. Large font shows known or predicted exon junctions. Exons are numbered for CES1, CES2 and CES5. Predicted transmembrane regions are shown.
* alternate residues for human CSII, CSE and CSEP. Residues involved in processing or N-glycanolation residues: [residue list] residues.

Charge clamp residues identified for human CSII: [residue list]. Seem known or predicted side functions. Sequences for cat, cow, and sheep are incomplete. Predicted phosphorylation sequence is also.
Figure 5: Carboxy-terminus nucleotide sequences and predicted amino acid sequences for dog and cat CES5 genes and proteins.

Sequences were derived from BLAT analyses using known sequences for dog and cat CES5 sequences and the UCSC web site (http://genome.ucsc.edu/cgi-bin/hgBlat?command=start) (see Table 1 for sources). * shows identical nucleotide residues following alignment of sequences. Stop indicates STOP codon for C-terminus observed in the cat CES7 gene. Note the apparent GT INDEL (insertion deletion) within the CES5 sequence generates an earlier termination of translation.

Dog CES5 C-terminus
TCAGATCACTTCGAGCCGGGAATACTTGTGGTTGTGGCTCTTTCGCTGTGGGCTCTGCTGGCTCTGCTGGCTCTGCTGGCTGTGGGGTCAGATCACTTCGAGCCGGGAATACTTGTGGTTGTGGCTCTTTCGCTGTGGGCTCTGCTGGCTCTGCTGGCTCTGCTGGCTGTGGGGTCAGATCACTTCGAGCCGGGAATACTTGTGGTTGTGGCTCTTTCGCTGTGGGCTCTGCTGGCTCTGCTGGCTCTGCTGGCT

Cat CES5 C-terminus
ATGGAATCACTTCGAGCCGGGAATACTTGTGGTTGTGGCTCTTTCGCTGTGGGCTCTGCTGGCTCTGCTGGCTCTGCTGGCTGTGGGGTCAGATCACTTCGAGCCGGGAATACTTGTGGTTGTGGCTCTTTCGCTGTGGGCTCTGCTGGCTCTGCTGGCTCTGCTGGCT

Figure 6: Phylogenetic tree of mammalian CES5 and of human CES1, CES2, CES3 and CES6 sequences.

The tree is labeled with the gene name and the species name. Note the separation of distinct CES1, CES2, CES3 and CES6 gene families from the mammalian CES5 family cluster. The gene duplication events generating five distinct gene families (CES1, CES2, CES3, CES5 and CES6) have been previously estimated to have occurred ~ 328-378 million years ago (Holmes et al., 2008).