Mammalian Carboxylesterase 3: Comparative Genomics and Proteomics

Roger S Holmes\textsuperscript{1,4}, Laura A Cox\textsuperscript{1,2} and John L VandeBerg\textsuperscript{1,2}

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

\textsuperscript{4}Corresponding 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; evolution; gene duplication.

Running Head: Mammalian CES3: comparative genomics and proteomics
Summary

At least five families of mammalian carboxylesterases (CES) catalyse the hydrolysis or transesterification of a wide range of drugs and xenobiotics and may also participate in fatty acyl and cholesterol ester metabolism. In this study, \textit{in silico} methods were used to predict the amino acid sequences, secondary and tertiary structures, and gene locations for CES3 genes and encoded proteins using data from several mammalian genome projects. Mammalian CES3 genes were located within a CES gene cluster with CES2 and CES6 genes, usually containing 13 exons transcribed on the positive DNA strand. Evidence is reported for duplicated CES3 genes for the chimp and mouse genomes. Mammalian CES3 protein subunits shared 58-97\% sequence identity and exhibited sequence alignments and identities for key CES amino acid residues as well as extensive conservation of predicted secondary and tertiary structures with those previously reported for human CES1. The human genome project has previously reported CES3 mRNA isoform expression in several tissues, particularly in colon, trachea and in brain. Predicted human CES3 isoproteins were apparently derived from exon shuffling and are likely to be secreted extracellularly or retained within the cytoplasm. Mouse CES3-like transcripts were localized in specific regions of the mouse brain, including the cerebellum, and may play a role in the detoxification of drugs and xenobiotics in neural tissues and other tissues of the body. Phylogenetic analyses demonstrated the relationships and potential evolutionary origins of the mammalian CES3 family of genes which were related to but distinct from other mammalian CES gene families.
Introduction

At least five families of mammalian carboxylesterases (CES; E.C.3.1.1.1) have been reported (Holmes et al., 2008a,b; 2009a-d) including CES1, the major liver enzyme (Munger et al., 1991; Shibata et al. 1993; Ghosh 2000; Holmes et al., 2009a); CES2, the major intestinal enzyme (Langmann et al. 1997; Schewer et al. 1997; Holmes et al., 2009a); CES3, expressed in liver, colon and brain (Sanghani et al. 2004; Holmes et al., 2009a-c); CES5 (also called CES7 or cauxin), a major urinary protein of the domestic cat (Miyazaki et al. 2003; 2006; Holmes et al., 2008b); and CES6, a predicted CES-like enzyme in brain (Clark et al., 2003; Holmes et al., 2009d). These enzymes catalyse hydrolytic and transesterification reactions with xenobiotics, anticancer prodrugs, and narcotics (Satoh & Hosokawa 1998; 2006; Satoh et al. 2002; Ohtsuka et al. 2003; Redinbo & Potter 2005), and detoxify organophosphates and insecticides (Ahmad & Forgash 1976; Leinweber 1987). CES may also catalyse several cholesterol and fatty acid metabolic reactions (Ghosh 2000; Tsujita & Okuda 1993; Becker et al. 1994; Hosokawa et al. 2007; Diczfalusy et al. 2001), the conversion of lung alveolar surfactant (Ruppert et al. 2006) and assist with the assembly of low density lipoprotein particles in liver (Wang et al. 2007).

Structures for several human and animal CES genes have been determined, including human (Becker et al. 1994; Langmann 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). Predicted gene structures have also been described for other mammalian CES gene families, including CES1, CES2 and CES5 genes, which are localized in two contiguous CES gene clusters on human (chromosome 16), horse (chromosome 3), cow (chromosome 18) and opossum (chromosome 1) genomes (Holmes et al. 2008a; 2009a-c). Mammalian CES genes usually contain 12 to 14 exons of DNA encoding CES enzyme sequences which may undergo exon shuffling generating several CES isoproteins in each case (Thierry-Mieg and Thierry-Mieg, 2006). There are similarities for the exon boundaries for each of the mammalian CES genes (Holmes et al., 2008a,b; 2009a-d) as well as significant sequence identities, especially for key regions previously identified for the major human liver CES1 (Bencharit et al. 2003; 2006; Fleming et al. 2005). Three-dimensional structural analyses of human CES1 have identified three major ligand binding sites, including the promiscuous active site, ‘side door’ and ‘Z-site’, where substrates, fatty acids and cholesterol analogues respectively, are bound; and the ‘gate’, reportedly functioning in product release (Bencharit et al. 2003; 2006; Fleming et al. 2005).

This paper reports the predicted gene structures and amino acid sequences for several mammalian CES3 genes and proteins, the predicted secondary and tertiary structures for mammalian CES3 protein subunits, and the
structural, phylogenetic and evolutionary relationships for these genes and enzymes with those for other mammalian CES gene families.

**Methods**

*In silico mammalian CES 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](http://blast.ncbi.nlm.nih.gov/Blast.cgi)) (Altschul *et al.*, 1997). Protein BLAST analyses used mammalian CES amino acid sequences previously described (Table 1). Non-redundant protein sequence databases for several mammalian genomes were examined using the blastp algorithm, including human (*Homo sapiens*) (International Human Genome Sequencing Consortium, 2001); chimp (*Pan troglodytes*) (Chimpanzee Genome Analysis Consortium, 2005); orangutan (*Pongo abelii*) ([http://genome.wustl.edu](http://genome.wustl.edu)); the cow (*Bos Taurus*) (Bovine Genome Project, 2008); horse (*Equus caballus*) (Horse Genome Project, 2008); mouse (*Mus musculus*) (Mouse Sequencing Consortium, 2002); and opossum (*Monodelphis domestica*) (Mikkelsen *et al.*, 2007). 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 CES-like proteins. These records were derived from annotated genomic sequences using the gene prediction method: GNOMON and predicted sequences with high similarity scores for human CES6. Predicted CES3-like protein sequences were obtained in each case and subjected to *in silico* analyses of predicted protein and gene structures.

BLAT analyses were subsequently undertaken for each of the predicted CES3 amino acid sequences using the UC Santa Cruz web browser ([http://genome.ucsc.edu/cgi-bin/hgBlat](http://genome.ucsc.edu/cgi-bin/hgBlat)) (Kent *et al.*, 2003) with the default settings to obtain the predicted locations for each of the mammalian CES3 genes, including predicted exon boundary locations and gene sizes. BLAT analyses were similarly undertaken of CES1, CES2, CES5 and CES6-like human genes using previously reported sequences for encoded CES subunits in each case (see Table 1). Structures for human CES3 isoforms (splicing variants) were obtained using the AceView website ([http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/index.html?human](http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/index.html?human)) to examine predicted gene and protein structures using the human CES3 cDNA sequence (GenBank NM024922) to interrogate this database of human mRNA sequences (Thierry-Mieg and Thierry-Mieg, 2006).

BLAT (BLAST-Like Alignment Tool) *in silico* studies were undertaken using the UC Santa Cruz web browser ([http://genome.ucsc.edu/cgi-bin/hgBlat](http://genome.ucsc.edu/cgi-bin/hgBlat)) (Kent *et al.*, 2003) with the default settings. UniProtKB/Swiss-Prot
Database [http://au.expasy.org] and GenBank [http://www.ncbi.nlm.nih.gov/Genbank/] sequences for human, orangutan and mouse CES sequences (Table 1) were used to interrogate human, chimp, orangutan, horse, cow, mouse and opossum genome sequences. Gene locations, predicted gene structures and CES protein subunit sequences were observed for each CES examined for those regions showing identity with the respective mammalian CES gene products.

**Predicted Structures and Properties of Mammalian CES Subunits.**

Predicted secondary and tertiary structures for human and other mammalian CES-like subunits were obtained using the PSIPRED v2.5 web site tools provided by Brunel University [http://bioinf.cs.ucl.ac.uk/psipred/psiform.html] (McGuffin et al. 2000) and the SWISS MODEL web tools [http://swissmodel.expasy.org/], respectively (Guex & Peitsch 1997; Kopp & Schwede 2004). The reported tertiary structure for the rabbit CES1 4-piperidino-piperidine complex (Bencharit et al., 2003) served as the reference for the predicted human CES3 tertiary structure, with a modeling range of residues 34 to 553. Theoretical isoelectric points and molecular weights for mammalian CES3 subunits were obtained using Expasy web tools (http://au.expasy.org/tools/pi_tool.html). SignalP 3.0 web tools were used to predict the presence and location of signal peptide cleavage sites (http://www.cbs.dtu.dk/services/SignalP/) for each of the predicted mammalian CES sequences (Emanuelsson et al 2007). The NetNGlyc 1.0 Server was used to predict potential N-glycosylation sites for human CES and mammalian CES3 subunits (http://www.cbs.dtu.dk/services/NetNGlyc/). Predictions of subcellular locations for mammalian CES3 ‘like’ enzymes and for human CES1, CES2, CES5 and CES6 were conducted using PSORT 11 (http://psort.ims.u-tokyo.ac.jp/form2.html) (Horton & Nakai, 1997).

**Mouse Brain CES3 Expression**

The Allen Mouse Brain Atlas was interrogated for cellular transcription of the mouse CES3 gene using the GenBank BC061004 sequence (see Table 1) (Lein et al., 2007). Data for sagittal sections of mouse brain were examined and recorded for mouse CES3 transcripts using web tools available on the Allen Brain Atlas web site (http://www.brain-map.org).

**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/]. An alignment score was calculated for each aligned sequence by first calculating a pairwise score for every pair of sequences aligned. Alignment ambiguous regions, including the amino and carboxyl
termini, were excluded prior to phylogenetic analysis yielding alignments of 453 residues for comparisons of mammalian CES1, CES2, CES3, CES5 and CES6 sequences (Table 1). Pairwise scores were calculated using the number of identities in the best alignment divided by the number of residues compared. Scores were initially calculated as percent identity scores and were converted to distances by dividing by 100 and subtracting from 1.0 to give the number of differences per site. The extent of divergence for the mammalian CES3-like subunits, and the human CES1, CES2, CES5 and CES6 subunits were determined using the SIM-Alignment tool for Protein Sequences [http://au.expasy.org/tools/sim-prot.html] (Schwede et al. 2003).

Results and Discussion

Alignments of Human CES Subunits and Mammalian CES3 Amino Acid Sequences.

The deduced amino acid sequences for horse, cow and mouse CES3-like subunits are shown in Figure 1 together with the previously reported sequences for human CES1 (Munger et al. 1991; Shibata et al. 1993); human CES2 (Langmann et al. 1997; Schewer et al. 1997); human CES3 (Sanghani et al. 2004); human CES5 (Holmes et al., 2008b); and human CES6 (Holmes et al., 2009d) (Table 1). Alignments of the human and other mammalian CES3 subunits examined in this figure showed between 68-77% sequence identities, suggesting that these CES protein subunits are products of the same family of genes, whereas sequence alignments of mammalian CES3 subunits with human CES1, CES2, CES5 and CES6 subunits exhibited lower levels of sequence identities with human CES3 (47%, 50%, 48% and 45%, respectively), indicating that these are members of distinct CES families (Table 2).

The amino acid sequence for human CES3 contained 571 residues whereas other predicted mammalian CES3 subunits contained fewer amino acids: 570 residues for cow and horse CES3; 554 residues for mouse CES3; and 568 residues for mouse CES3L (Figure 1). Previous studies on human CES1 have identified key residues which contribute to the catalytic, subcellular localization, oligomeric and regulatory functions for this enzyme (sequence numbers refer to human CES1). These include the catalytic triad for the active site (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); 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. Identical residues were observed for each of the mammalian CES3 subunits for the active site triad and disulfide bond forming residues for human CES1 although several changes were
observed for the corresponding residues for CES3 subunits of other key human CES1 residues: the ‘side-door’ identified for human CES1 (Val424-Met425-Phe426) (Bencharit et al., 2003; 2006) has a Val416-Phe417-Ile418 sequence for human CES3; the ‘Z-site’ (Gly358 for human CES1) has been changed to Ser355 for human CES3, although other mammalian CES3 sequences examined have retained a glycine residue at this position; the hydrophobic N-terminal sequence for mammalian CES3 sequences has undergone major changes however this region retains in each case a predicted signal peptide property; the mammalian CES3 C-terminal sequences have undergone significant changes in sequence (QEDL, QEEL or PEEL) in comparison with human CES1 (HIEL). This may be responsible for a change in the predicted microlocalization for the CES3 subunits (secreted or extracellular enzymes), as compared with human CES1, which is retained by the endoplasmic reticulum (Table 1). The ‘gate’ residue has also undergone changes for the mammalian CES3 sequences examined from Phe551 (CES1) to Tyr (human and horse CES3) or Leu (cow and mouse CES3).

Other key human CES1 sequences included two charge clamps which are responsible for subunit-subunit interaction, namely residues Lys78/Glu183 and Glu72/Arg193 (Bencharit et al. 2003; 2006; Fleming et al. 2005). Predicted mammalian CES3 subunit sequences for these sites indicated that only one of the charge clamps would be retained for human CES3 subunits: Arg88/Glu191 (Figure 1), whereas other mammalian CES3 subunits examined showed no potential for forming charge clamps at the corresponding sequence positions. Previous reports have shown that human and baboon CES2 subunits, which lack these charge clamp residues, behave as monomers (Pindel et al. 1997; Holmes et al., 2009a). With the exception of human CES3, the mammalian CES3 sequences have undergone amino acid substitutions for those residues which contribute to the oligomeric subunit structure for human CES1. The identified human CES1 charge clamp residues Lys78/Glu183 and Glu72/Arg193 have other residues at the corresponding sites (Arg88/191Lys and Gln82/Ala194 for the horse CES3 sequence) (Figure 1), and it is likely that these mammalian CES3 subunits are also monomeric in subunit structure. This may have a major influence on the kinetics and biochemical roles for mammalian CES3 since three dimensional studies have suggested that ligand binding to the human CES1 ‘Z-site’ shifts the trimer-hexamer equilibrium towards the trimer facilitating substrate binding and enzyme catalysis (Redinbo & Potter 2005). This property is unlikely to be shared by mammalian CES3 which is expected to display distinct kinetic properties to the oligomeric CES1 enzyme.

The N-glycosylation site reported for human CES1 at Asn79-Ala80-Thr81 (Kroetz et al. 1993; Bencharit et al. 2003; 2006; Fleming et al. 2005) has not been retained for any of the mammalian CES3 sequences, although additional potential N-glycosylation sites were observed at one or several positions (Table 3). Given the reported role of the N-glycosylated carbohydrate group in contributing to mammalian CES stability and maintaining catalytic
efficiency (Kroetz et al. 1993), this property may be shared by the mammalian CES3 subunits as well, especially for those containing multiple potential sites for N-glycosylation, such as the mouse CES3L sequence, which contains four such sites.

Predicted Secondary and Tertiary Structures for Mammalian CES3 Subunits.

Analyses of predicted secondary structures for mammalian CES3 sequences were compared with the previously reported secondary structure for human CES1 (Bencharit et al., 2003; 2006) and with those predicted for human CES2, CES5 and CES6 (Holmes et al., 2009a) (Figure 1). Similar α-helix β-sheet structures were observed for all of the mammalian CES3 and human CES gene products examined. Consistent structures were apparent near key residues or functional domains including the α-helix within the N-terminal signal peptide; the β-sheet and α-helix structures near the active site Ser228 (human CES1) and ‘Z-site’ (Glu354/Gly356 respectively); the α-helices bordering the ‘side door’ site; and the α-helix containing the ‘gate’ residue (Phe551 for human CES1). In addition, two random coil regions (residues 51-115 and 169-188 for human CES1) were predominantly retained for all forms of mammalian CES3 examined which have been shown for human CES1 to contain 2 charge clamps sites: (Lys79/Glu183 and Glu73/Arg186); an N-glycosylation site at Asn79-Ala80-Thr81; a second potential N-glycosylation site for human CES2, and one of the disulfide bridges (87Cys/117Cys) reported for human CES1. The human CES5 secondary structure, however, predicted an additional helix at the hydrophobic C-termini, in comparison with mammalian CES3 C-terminal sequences. In addition, the predicted 3-D structure for human CES3 is similar to the reported for human CES1 (Bencharit et al., 2003; 2006) (Figure 3). These were based on structures reported for residues 21-553 of human CES1 and a predicted structure for residues 34-553 of human CES3. The single helix and following two β-sheet structures at the C-termini (shown in red at the top) for the human CES1 and CES3 subunits are readily apparent, as are four β-sheet structures at the N-termini of the CES subunits (shown in blue at the base), in each case. It is apparent that human CES3 is likely to be highly similar in its secondary and tertiary structure to that reported for human CES1.

Predicted Gene Locations and Exonic Structures for Mammalian CES3 Genes.

Table 1 summarizes the predicted locations for mammalian CES3 genes based upon BLAT interrogations of several mammalian genomes using the reported sequences for human CES3, mouse CES3 and CES3L, and the predicted sequences for horse and cow CES3 and the UC Santa Cruz Web Browser (Kent et al. 2003). The predicted mammalian CES3 genes were located in a CES2/CES3/CES6 gene cluster in each case, although the gene order underwent changes for different species. In addition, mouse showed evidence of duplicate CES3 genes, with
predicted CES3 and CES3L genes being located within a four CES gene cluster on chromosome 8: CES2-CES3-CES3L-CES6 (Table 1). Chimp CES3 also provided evidence for duplicate forms of this gene on chromosome 16 but with a changed order of genes: CES6-CES3L-CES2-CES3 (Table 1). BLAT interrogations of mammalian genomes with the corresponding CES sequences suggested that the CES2-CES3-CES6 gene cluster was syntenic for chromosomes 16 (human, chimp and orangutan), 3 (horse), 18 (cow), 8 (mouse) and 1 (opossum), respectively.

With the exception of the opossum CES3 gene, mammalian CES3 ‘like’ genes were apparently transcribed on the positive strand, as for the human CES2 and CES6 genes. In contrast, human CES1 and CES5 genes were transcribed on the negative strand (Table 1). Figure 1 summarizes the predicted exonic start sites for cow, horse and mouse CES3 genes with each having 13 exons, in identical or similar positions to those predicted for the human CES3 gene (Holmes et al., 2009a). The predicted chimp CES3L gene however exhibited only 11 predicted exons.

Human CES1 (Munger et al. 1991; Shibata et al. 1993), CES2 (Tang et al., 2008), CES5 (Ota et al., 2004) and CES6 (Holmes et al., 2009d) genes contained 14, 12, 13 and 14 exons respectively (Figure 1), in similar positions to those predicted for the mammalian CES3 genes (Figure 1).

Figure 4 shows the predicted nucleotide sequences for exons 2-4 and introns 2 and 3 of mouse CES3 and CES3L genes. The sequences for the exons were 98% identical with nearly identical amino acid sequences involving five amino acid substitutions for these genes, whereas mouse CES3 and CES3L introns 2 and 3 showed distinct nucleotide sequences which were 85% and 94% identical, respectively. In addition, BLAT analyses of the mouse genome revealed distinct but closely localized predicted positions for these genes, being separated by ~26 kilobases of DNA on chromosome 8 (Table 1). These data support an hypothesis for two closely related and localized CES3-like genes in the mouse, for which the subunits have been previously designated as esterase-31 (also called CES3) (Aida et al., 1993) and esterase-31L (also called CES3L) (Mouse Genome Consortium, 2004).

**Mouse Brain CES3 and CES3L Expression.**

Figure 5 shows a sagittal section of mouse brain examining the distribution of CES3 and CES3L transcripts (GenBank mRNA BC061004 and BC019047) within different regions of the brain (provided by the Allen Institute for Brain Science) (Lein et al., 2007). The cerebellum folds exhibited highest staining levels for both transcripts although staining was observed throughout the brain, including the hippocampus, amygdalar nuclei, the olfactory bulb, the cortex, the pons and the medulla regions.

The differential tissue and subcellular distributions of CES family members may assist in determining the differential roles of CES family members within mammalian organisms. Mammalian liver is predominantly...
responsible for drug and xenobiotic clearance from the body with CES1 and CES2 (with CES1 > CES2) playing major roles, following absorption of drugs and xenobiotics into the circulation (Pindel et al. 1997; Imai 2006).

Mammalian intestine (with CES2 > CES1) is predominantly responsible for first pass clearance of several drugs and xenobiotics, with the activity occurring mostly in the ileum and jejunum and processed via CES2 (Imai et al. 2003). CES1 and CES2 also serve different roles in prodrug activation, as shown for the anti-cancer drug irinotecan (CPT-11) which is converted to its active form SN-38 predominantly by CES2 (Humerickhouse et al. 2000; Xu et al. 2002). In contrast with CES1 and CES2, mammalian CES5 is predominantly expressed in peripheral tissues, including brain, kidney, lung and testis (Thierry-Mieg and Thierry-Mieg, 2006), and in cat, is a secreted form of CES enzyme appearing in cat urine (Miyazaki et al. 2003; 2006). Mammalian CES5 may serve in two major roles within mammalian fluids and peripheral tissues, including regulating the production of a pheromone precursor in urine (Miyazaki et al. 2003; 2006) and contributing to lipid and cholesterol transfer processes within male reproductive fluids (Ecroyd et al. 2005).

The metabolic role(s) for mammalian CES3 has not been extensively investigated although the enzyme is capable of activating prodrugs such as irinotecan (Sanghani et al., 2004) and the gene is expressed as distinct isoforms in tissues of the body, including the colon, placenta and neural tissues, such as the cerebellum and hippocampus (Figure 2; Table 4) (Thierry-Mieg and Thierry-Mieg, 2006). Human CES3 mRNAs encode isoforms which are predicted to be either secreted or extracellular (CES3a-c), cytosolic (CES3d) or extracellular (CES3e-u), and may play distinct roles in carboxyl-ester metabolism in the body. A comparison of the predicted cellular locations for mammalian CES3 reveals that it is likely to be extracellular (including the cell wall) in primate (human, chimp and orangutan) and mouse tissues, whereas horse and bovine CES3 are predicted to be localized in the endoplasmic reticulum, which is comparable with human CES1 and CES2. The extracellular locations predicted for human and mouse CES3 may reflect distinct roles for this enzyme in drug metabolism within peripheral tissues of the body.

**Phylogeny and Divergence of CES3 and Other Mammalian CES Sequences.**

A phylogenetic tree (Figure 6) was calculated by the progressive alignment of human, chimp, orangutan, horse, cow, mouse and opossum CES3-like amino acid sequences with human CES1, CES2, CES5 and CES6 sequences. The phylogram showed clustering of all of the CES3-like sequences which were distinct from the other human CES families. In addition, the CES3 and CES3L chimp and mouse sequences showed clustering with the primate and mouse CES3-like genes, respectively which is consistent with these genes being products of recent
duplication events during mammalian evolution. Holmes et al (2008a) have previously demonstrated that the gene duplication events which generated the ancestral mammalian CES1, CES2, CES3, CES5 and CES6 genes predated the common ancestor for marsupial and eutherian mammals [estimated at ~ 173-193 million years ago by Woodburne et al (2003)], and appear to coincide with the early diversification of tetrapods approximately 350-360 million years ago ((Donoghue & Benton, 2007). The mammalian CES3 family of genes may therefore be considered to be of an ancient origin, well prior to the appearance of mammals in evolutionary time.

In conclusion, the results of the present study indicate that mammalian CES3 genes and encoded CES3 enzymes represent a distinct CES gene and enzyme family which share key conserved sequences and structures that have been reported for human CES1 and have family specific sequences consistent with the oligomeric and monomeric subunit structures for CES1 and CES3, respectively. This study also reports that chimp and mouse genomes contain at least two CES3 ‘like’ genes, which are located in tandem with the CES2 and CES6 genes, and with the more distantly located CES1 and CES5 genes on chromosomes 16 and 8, respectively. Predicted secondary structures and tertiary structures for mammalian CES3 subunits showed a high degree of conservation with human CES1. Phylogeny studies using several mammalian CES3 subunits (human, chimp, orangutan, mouse, horse, cow and opossum) indicated that CES3 ‘like’ genes have apparently appeared during mammalian evolution well prior to the eutherian and marsupial common ancestor more than 170 million years ago, which is consistent with previous studies (Holmes et al, 2008a).

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


Bovine Genome Project (2008) http://hgsc.bcm.tmc.edu/projects/bovine


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.


Figure 1: Amino Acid Sequence Alignments for Mammalian CES3 and Human CES1, CES2, CES5 and CES6 Subunits

See Table 1 for sources of CES sequences; * shows identical residues for CES subunits; : similar alternate residues; . dissimilar alternate residues; Residues involved in endoplasmic reticulum processing at N- (Signal peptide) and C- termini (MTS-microsomal (endoplasmic reticulum) targeting sequence); N-glycosylation sites at 79NAT (Human CES1) and potential N-glycosylation sites; Active site (AS) triad residues Ser; Glu; and His. 'Gate' residues and Cholesterol binding Gly residue (Z site) for human CES. 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. Bold font shows known or predicted exon junctions. Exon numbers refer to human CES1 gene.
Figure 2: Gene Structures and Splicing Variants for the Human CES3 (NM024922) and the Mouse CES3 (BC061004) Genes.

Derived from AceView website (Thierry-Mieg and Thierry-Mieg, 2006)
Mature isoform variants (a,b,c etc) are shown with capped 5’- and 3’- ends for the predicted mRNA sequences. Gene COesterase 1 refers to human CES3 gene; Gene Es31 refers to mouse CES3 gene.
Figure 3: Tertiary Structure for Human CES1 Subunit and Predicted Tertiary Structure for Human CES3 Subunit.

The structure for human CES1 is taken from Bencharit et al, 2003. Predicted human CES3 3-D structure was obtained using the SWISS MODEL web site [http://swissmodel.expasy.org/workspace/index.php](http://swissmodel.expasy.org/workspace/index.php) and the reported amino acid sequence (Sanghani et al., 2004). The rainbow color code describes the 3-D structures from the N- (blue) to C-termini (red color).
Figure 4: Nucleotide Sequence Alignments for Mouse CES3 ‘Like’ Genes: CES3 (Es31) and CES3L (Es31L): Predicted Exons 2–4; and Predicted Introns 2 and 3.

Exon sequences are in CAPITALS and intron sequence in lower case. Note the different percentages of sequence identities for exon 2–4, (98%); introns 2–3 (92%). Nucleotide substitutions are shaded. The deduced amino acid sequence for exons 2–4 of mouse CES3 is shown as well as any deduced amino acid substitutions observed for CES3L.
Figure 5: Sagittal section of mouse brain showing the distribution of mouse CES3 and CES3L transcripts.

The two distinct regions of the mouse brain (C57BL6/6J strain) included the cerebellum (right) and the cerebrum (left). Note major staining of CES3 and CES3L transcripts in the cerebellum folds and the hippocampus and amygdalar nuclei regions of the cerebrum. The sections were obtained from the Allen Brain Atlas web site (http://www.brain-map.org) (Lein et al., 2007) using the GenBank IDs BC061005 and BC019147, respectively.
Figure 6: Phylogenetic Tree of Mammalian CES3 and Human CES1, CES2, CES5 and CES6 Amino Acid Sequences.

The tree is labeled with the gene name and the name of the mammal. Note the major cluster for mammalian CES3 'like' sequences and the separation of these sequences from human CES1, CES2, CES5 and CES6 sequences. The gene duplication events generating the 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., 2008a).
<table>
<thead>
<tr>
<th>Mammal</th>
<th>CES Gene</th>
<th>GenBank ID</th>
<th>UNIPROT Accession No.</th>
<th>No of Amino Acids</th>
<th>Chromosome Location</th>
<th>Strand</th>
<th>*CES Gene order</th>
<th>Gene Size (bps)</th>
<th>No. of Exons</th>
<th>*pI</th>
<th>Subunit Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chimp</td>
<td>CES3</td>
<td>*XP_511224</td>
<td></td>
<td>571</td>
<td>16: 5,892,348-5,905,878 16: 54,242,602-54,266,316</td>
<td>Positive</td>
<td>CES6-CES3L-CES2-CES3</td>
<td>13,879</td>
<td>13</td>
<td>5.7</td>
<td>extracellular</td>
</tr>
<tr>
<td>Chimpanzee</td>
<td>CES3L</td>
<td>*XP_001173132</td>
<td></td>
<td>479</td>
<td>16: 65,552,639-65,566,552</td>
<td>Positive</td>
<td>CES6-CES3-CES3</td>
<td>13,531</td>
<td>11</td>
<td>6.6</td>
<td>extracellular</td>
</tr>
<tr>
<td>Orangutan</td>
<td>CES3</td>
<td>CR857194</td>
<td>Q5RCL7</td>
<td>569</td>
<td>3: 16,671,230-33,683,956</td>
<td>Positive</td>
<td>CES2-CES3-CES6</td>
<td>10,434</td>
<td>13</td>
<td>6.3</td>
<td>extracellular</td>
</tr>
<tr>
<td>Horse</td>
<td>CES3</td>
<td>*XP_001496251</td>
<td></td>
<td>570</td>
<td>8: 107,572,572-107,582,000</td>
<td>Positive</td>
<td>CES2-CES3-CES6</td>
<td>9,429</td>
<td>13</td>
<td>5.9</td>
<td>extracellular</td>
</tr>
<tr>
<td>Cow</td>
<td>CES3</td>
<td>*XP_590749</td>
<td></td>
<td>570</td>
<td>8: 107,607,684-107,617,159</td>
<td>Positive</td>
<td>CES2-CES3-CES6</td>
<td>9,476</td>
<td>13</td>
<td>5.7</td>
<td>extracellular</td>
</tr>
<tr>
<td>Mouse</td>
<td>CES3</td>
<td>BC061004</td>
<td>Q63880</td>
<td>554</td>
<td>1: 677,687,773-677,741,087</td>
<td>Negative</td>
<td>^CES2-CES3-L-CES6</td>
<td>53,315</td>
<td>13</td>
<td>5.6</td>
<td>extracellular</td>
</tr>
<tr>
<td>Mouse</td>
<td>CES3L</td>
<td>BC019147</td>
<td></td>
<td>568</td>
<td>1: 677,687,773-677,741,087</td>
<td>Negative</td>
<td>^CES2-CES3-L-CES6</td>
<td>53,315</td>
<td>13</td>
<td>5.6</td>
<td>extracellular</td>
</tr>
</tbody>
</table>

Table 1: Mammalian CES3 and Human CES1, CES2, CES5 and CES6 Genes and Subunits.

*RefSeq NCBI Sequence; ^the mouse and opossum genomes contain multiple contiguous CES2 genes; extracellular and/or cell wall location; ER: endoplasmic reticulum location; *pI refers to predicted isoelectric point. *CES gene in bold refers to the specific gene among the contiguous CES genes.
Table 2: Percentage identities for mammalian CES3 and human CES1, CES2, CES5 and CES6 subunit amino acid sequences.

<table>
<thead>
<tr>
<th>CES Gene</th>
<th>Human CES1</th>
<th>Chimpanzee CES1</th>
<th>Orangutan CES1</th>
<th>Horse CES1</th>
<th>Cow CES1</th>
<th>Mouse CES1</th>
<th>Opossum CES1</th>
<th>Human CES2</th>
<th>Human CES5</th>
<th>Human CES6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human CES3</td>
<td>100</td>
<td>97</td>
<td>94</td>
<td>92</td>
<td>77</td>
<td>73</td>
<td>68</td>
<td>61</td>
<td>58</td>
<td>47</td>
</tr>
<tr>
<td>Chimpanzee CES3</td>
<td>97</td>
<td>100</td>
<td>95</td>
<td>92</td>
<td>78</td>
<td>75</td>
<td>61</td>
<td>61</td>
<td>58</td>
<td>47</td>
</tr>
<tr>
<td>Orangutan CES3</td>
<td>94</td>
<td>95</td>
<td>100</td>
<td>93</td>
<td>78</td>
<td>74</td>
<td>61</td>
<td>61</td>
<td>58</td>
<td>47</td>
</tr>
<tr>
<td>Chimpanzee CES3L</td>
<td>92</td>
<td>92</td>
<td>90</td>
<td>100</td>
<td>76</td>
<td>73</td>
<td>67</td>
<td>67</td>
<td>58</td>
<td>46</td>
</tr>
<tr>
<td>Horse CES3</td>
<td>77</td>
<td>70</td>
<td>76</td>
<td>76</td>
<td>100</td>
<td>79</td>
<td>61</td>
<td>61</td>
<td>58</td>
<td>47</td>
</tr>
<tr>
<td>Cow CES3</td>
<td>70</td>
<td>75</td>
<td>74</td>
<td>73</td>
<td>79</td>
<td>100</td>
<td>61</td>
<td>61</td>
<td>47</td>
<td>47</td>
</tr>
<tr>
<td>Mouse CES3</td>
<td>68</td>
<td>68</td>
<td>68</td>
<td>67</td>
<td>66</td>
<td>100</td>
<td>91</td>
<td>57</td>
<td>45</td>
<td>45</td>
</tr>
<tr>
<td>Mouse CES3L</td>
<td>68</td>
<td>68</td>
<td>68</td>
<td>67</td>
<td>65</td>
<td>91</td>
<td>100</td>
<td>56</td>
<td>45</td>
<td>47</td>
</tr>
<tr>
<td>Opossum CES3</td>
<td>58</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>51</td>
<td>57</td>
<td>51</td>
<td>100</td>
<td>49</td>
<td>49</td>
</tr>
<tr>
<td>Human CES1</td>
<td>47</td>
<td>47</td>
<td>47</td>
<td>46</td>
<td>49</td>
<td>47</td>
<td>45</td>
<td>45</td>
<td>43</td>
<td>100</td>
</tr>
<tr>
<td>Human CES2</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>49</td>
<td>50</td>
<td>47</td>
<td>43</td>
<td>47</td>
<td>45</td>
<td>51</td>
</tr>
<tr>
<td>Human CES5</td>
<td>48</td>
<td>48</td>
<td>48</td>
<td>49</td>
<td>47</td>
<td>43</td>
<td>47</td>
<td>43</td>
<td>45</td>
<td>49</td>
</tr>
<tr>
<td>Human CES6</td>
<td>45</td>
<td>44</td>
<td>44</td>
<td>43</td>
<td>41</td>
<td>43</td>
<td>41</td>
<td>41</td>
<td>45</td>
<td>46</td>
</tr>
</tbody>
</table>

Table 2: Percentage identities for mammalian CES3 and human CES1, CES2, CES5 and CES6 subunit amino acid sequences.
Table 3: Predicted N-glycosylation sites for mammalian CES3 and human CES1, CES2, CES5 and CES6 subunits.

The identified N-glycosylation site is for human CES1 (Kroetz et al., 1993). Amino acid residues are shown for potential N-glycosylation sites: N-Asn; A-Ala; T-Thr; S-Ser; M-Met; L-Leu; D-Asp; G-Gly; P-Phe; I-Ile.

<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>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>Site 15</th>
<th>No of Sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>CES1</td>
<td>70NAT</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>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CES2</td>
<td>110NMT</td>
<td>274NLS</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>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CES5</td>
<td>286NAS</td>
<td>339NLS</td>
<td>313NLT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CES6</td>
<td>244NVT</td>
<td>276NT</td>
<td>370NIT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Human</td>
<td>CES3</td>
<td>10NASS</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>1</td>
<td></td>
</tr>
<tr>
<td>Cow</td>
<td>CES1</td>
<td>100NAS</td>
<td>212NT</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>2</td>
<td></td>
</tr>
<tr>
<td>Horse</td>
<td>CES1</td>
<td>212NT</td>
<td>244NAS</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>3</td>
<td></td>
</tr>
<tr>
<td>Mouse</td>
<td>CES1</td>
<td>298NAS</td>
<td>309NDS</td>
<td>389NNS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Mouse</td>
<td>CES3</td>
<td>300NAS</td>
<td>326NDS</td>
<td>389NNS</td>
<td>423NFS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>N-signal (location)</td>
<td>Peptide Subcellular Location</td>
<td>Tissue Distribution (mRNA Frequency)</td>
<td>Predicted CES type B Domain</td>
<td>GenBank ID</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>---------------------</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></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>yes (26-27)</td>
<td>Secreted or extracellular</td>
<td>colon (8), breast, trachea (4) brain, placenta, rectum (2)</td>
<td>yes NM024922</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></td>
</tr>
<tr>
<td>yes (26-27)</td>
<td>Secreted or extracellular</td>
<td>colon (10) trachea (6) brain, breast (5) placenta (4) testis (3)</td>
<td>yes BC053670</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></td>
</tr>
<tr>
<td>yes (26-27)</td>
<td>Secreted or extracellular</td>
<td>colon (10) trachea (6) brain breast (5) placenta (4) testis (3)</td>
<td>yes CR618303</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></td>
</tr>
<tr>
<td>yes (26-27)</td>
<td>Cytosol</td>
<td>ovary, teratocarcinoma, stomach, trachea (1)</td>
<td>yes *DB232966</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></td>
</tr>
<tr>
<td>yes (26-27)</td>
<td>Extracellular</td>
<td>colon (5), nerve tumor, uterus, brain (2)</td>
<td>yes AK025389</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></td>
</tr>
<tr>
<td>yes (26-27)</td>
<td>Extracellular</td>
<td>bladder, uterus (1)</td>
<td>yes DA040317</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></td>
</tr>
<tr>
<td>yes (26-27)</td>
<td>Extracellular</td>
<td>trachea (1)</td>
<td>yes DB215224</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></td>
</tr>
<tr>
<td>no</td>
<td>Extracellular</td>
<td>stomach (1)</td>
<td>no AW813514</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></td>
</tr>
</tbody>
</table>

**Table 4: Tissue distribution of human mRNA transcripts and predicted encoded CES3 isoforms.**

Predicted CES3 gene sizes (kilobases); CES type B domain refers to the presence of essential amino acids at the CES active site; N-signal peptide location refers to the predicted site of hydrolysis for the N-terminal peptide involved in intracellular CES3 processing. Note the higher frequencies for CES3a, CES3b and CES3c expressed in colon, trachea and breast tissue. * refers to a partial mRNA sequence.