Published in final edited form as:
Mamm Genome. 2010 October ; 21(9-10): 427–441. doi:10.1007/s00335-010-9284-4.

Recommended nomenclature for five mammalian carboxylesterase gene families: human, mouse, and rat genes and proteins

Roger S. Holmes
Department of Genetics, Southwest Foundation for Biomedical Research, San Antonio, TX 78227-5301, USA

Southwest National Primate Research Center, Southwest Foundation for Biomedical Research, San Antonio, TX, USA

School of Biomolecular and Physical Sciences, Griffith University, Brisbane, QLD, Australia

Matthew W. Wright
European Bioinformatics Institute, Wellcome Trust Genome Campus, Cambridge, UK

Stanley J. F. Laulederkind
Rat Genome Database, Human Molecular Genetics Center, Medical College of Wisconsin, Milwaukee, WI, USA

Laura A. Cox
Department of Genetics, Southwest Foundation for Biomedical Research, San Antonio, TX 78227-5301, USA

Southwest National Primate Research Center, Southwest Foundation for Biomedical Research, San Antonio, TX, USA

Masakiyo Hosokawa
Laboratory of Drug Metabolism and Biopharmaceutics, Chiba Institute of Science, Choshi, Chiba, Japan

Teruko Imai
Graduate School of Pharmaceutical Sciences, Kumamoto University, Kumamoto, Japan

Shun Ishibashi
Department of Medicine, Jichi Medical University, Shimotsuke, Tochigi, Japan

Richard Lehner
CIHR Group on Molecular and Cell Biology of Lipids, University of Alberta, Edmonton, AB, Canada

Masao Miyazaki
The Institute of Glycoscience, Tokai University, Kanagawa, Japan

Everett J. Perkins
Department of Drug Disposition, Lilly Research Laboratories, Eli Lilly and Company, Indianapolis, IN, USA

Phillip M. Potter

rholmes@sfbrgenetics.org .

Department of Molecular Pharmacology, St. Jude Children's Research Hospital, Memphis, TN, USA

Matthew R. Redinbo
Department of Chemistry, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

Jacques Robert
Laboratoire de Pharmacologie, Institut Bergonié, Bordeaux Cedex, France

Tetsuo Satoh
Graduate School of Pharmaceutical Sciences, Chiba University, Chiba, Japan

Tetsuro Yamashita
Department of Agro-bioscience, Iwate University, Morioka, Japan

Bingfan Yan
Department of Biomedical and Pharmaceutical Sciences, University of Rhode Island, Kingston, RI, USA

Tsuyoshi Yokoi
Division of Pharmaceutical Sciences, Graduate School of Medical Science, Kanazawa University, Kanazawa, Japan

Rudolf Zechner
Institute of Molecular Biosciences, University of Graz, Graz, Austria

Lois J. Maltais
The Jackson Laboratory, Bar Harbor, ME, USA

Abstract

Mammalian carboxylesterase (*CES* or *Ces*) genes encode enzymes that participate in xenobiotic, drug, and lipid metabolism in the body and are members of at least five gene families. Tandem duplications have added more genes for some families, particularly for mouse and rat genomes, which has caused confusion in naming rodent *Ces* genes. This article describes a new nomenclature system for human, mouse, and rat carboxylesterase genes that identifies homolog gene families and allocates a unique name for each gene. The guidelines of human, mouse, and rat gene nomenclature committees were followed and "*CES*" (human) and "*Ces*" (mouse and rat) root symbols were used followed by the family number (e.g., human *CES1*). Where multiple genes were identified for a family or where a clash occurred with an existing gene name, a letter was added (e.g., human *CES4A*; mouse and rat *Ces1a*) that reflected gene relatedness among rodent species (e.g., mouse and rat *Ces1a*). Pseudogenes were named by adding "*P*" and a number to the human gene name (e.g., human *CES1P1*) or by using a new letter followed by *ps* for mouse and rat *Ces* pseudogenes (e.g., *Ces2d-ps*). Gene transcript isoforms were named by adding the GenBank accession ID to the gene symbol (e.g., human *CES1_AB119995* or mouse *Ces1e_BC019208*). This nomenclature improves our understanding of human, mouse, and rat *CES*/*Ces* gene families and facilitates research into the structure, function, and evolution of these gene families. It also serves as a model for naming *CES* genes from other mammalian species.

Introduction

Five families of mammalian carboxylesterases (CES; E.C.3.1.1.1) have been described, including CES1, the major liver enzyme (Ghosh 2000; Holmes et al. 2009a; Munger et al. 1991; Shibita et al. 1993); CES2, the major intestinal enzyme (Holmes et al. 2009a; Langmann et al. 1997; Schewer et al. 1997); CES3, expressed in brain, liver, and colon (Holmes et al. 2010; Sanghani et al. 2004); CES5 (also called CES7 or cauxin), a major

urinary protein of the domestic cat also present in human tissues (Holmes et al. 2008a; Miyazaki et al. 2003, 2006; Zhang et al. 2009); and CES6, a predicted CES-like enzyme in brain (Clark et al. 2003; Holmes et al. 2009a; reviewed by Williams et al. 2010). These enzymes catalyze hydrolytic and transesterification reactions with xenobiotics, anticancer prodrugs, and narcotics (Ohtsuka et al. 2003; Redinbo and Potter 2005; Satoh and Hosokawa 1998, 2006; Satoh et al. 2002), the conversion of lung alveolar surfactant (Ruppert et al. 2006), and several lipid metabolic reactions (Becker et al. 1994; Diczfalusy et al. 2001; Ghosh 2000; Hosokawa et al. 2007; Tsujita and Okuda 1993); they may also assist with the assembly of low-density lipoprotein particles in liver (Wang et al. 2007).

Structures for human and animal *CES* genes have been reported, including rodent *CES1*- and *CES2*-"like" genes (Dolinsky et al. 2001; Ghosh et al. 1995; Hosokawa et al. 2007) and human *CES1* and *CES2* genes (Becker et al. 1994; Ghosh 2000; Langmann et al. 1997; Marsh et al. 2004). Predicted gene structures have been also described for the human *CES3*, *CES5*, and *CES6* genes, which are localized with *CES1* and *CES2* in two contiguous *CES* gene clusters on human chromosome 16 (Holmes et al. 2008a, 2009a, b, 2010). In addition, a *CES1*-like pseudogene (currently designated *CES4*) is located with the *CES1–CES5* gene cluster (Yan et al. 1999). Mammalian *CES* genes usually contain 12–14 exons of DNA encoding CES enzyme sequences which may be shuffled during mRNA synthesis, generating several *CES* transcripts and enzymes encoded by each of the *CES* genes (see Thierry-Mieg and Thierry-Mieg 2006). There are significant sequence similarities for the five *CES* families, especially for key regions previously identified for 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 broad-specificity active site, the "side door," and the "Z-site," where substrates, fatty acids, and cholesterol analogs, respectively, are bound; and an active site `gate', which may facilitate product release following catalysis (Bencharit et al. 2003, 2006; Fleming et al. 2005).

Because of the confusion associated with the current nomenclature for mammalian *CES* genes, particularly for mouse and rat *Ces* genes where significant gene duplication events have generated a large number of *Ces1*-like and *Ces2*-like genes (Berning et al. 1985; Dolinsky et al. 2001; Ghosh et al. 1995; Hosokawa et al. 2007; Satoh and Hosokawa 1995), this article proposes a new nomenclature system that enables easy identification of *CES* family members for this enzyme. The nomenclature follows the guidelines of the human, mouse, and rat gene nomenclature committees and allocates a new name for each human (*CES*) or mouse and rat (*Ces*) gene. It also names and identifies the gene family origin for identified *CES* pseudogenes and provides a system for naming transcript iso-forms derived from each of the *CES* genes. The nomenclature has the flexibility to accommodate new human, mouse, and rat *CES* genes and will assist further research into the structure, function, and evolution of these gene families as well as serve as a model for naming *CES* genes from other mammalian species.

Guiding principles for the new *CES* nomenclature

The new nomenclature system for human, mouse, and rat *CES* genes and enzymes is based on the identification of homolog gene families and a subsequent allocation of a unique gene name for each of the genes observed from genome databases or reported from previous studies. It follows the guidelines of the human, mouse, and rat gene nomenclature committees and recommends the naming of homolog *CES* or *Ces* genes among species. The italicized root symbol "*CES*" for human and "*Ces*" for mouse and rat genes were used, followed by an number describing the gene family (examples include *CES1* for human *CES* family 1 or *Ces1* for mouse and rat *Ces* family 1 genes) (Tables 1, 2, 3). For mammalian genomes in which multiple genes were identified or a gene required a name that clashed

with an existing name, a capital letter (for human genes) (e.g., *CES4A*) or a lower-case letter (for mouse and rat genes) (e.g., *Ces1a*, *Ces1b* for multiple mouse *Ces1*-like genes) was added after the number. The letter used for multiple genes reflected the relatedness of the genes across species (e.g., reflecting higher degrees of identity for mouse and rat *Ces1a* genes). When a human *CES* pseudogene was identified, a capital "P" and a number were added to the gene name (e.g., *CES1P1*), whereas for mouse and rat *Ces* pseudogenes, a unique lower-case letter was used followed by "-ps" (e.g., *Ces2d-ps*). Transcript iso-forms of human (*CES*) and mouse and rat (*Ces*) gene transcripts were designated by following the gene name with the GenBank transcript ID, such as human *CES1_AB119997* and *CES1_AB187225*, which differs from the current nomenclature used for human *CES1* iso-forms (*CES1A1* and *CES1A2*, respectively) (see Table 1).

Human *CES* genes

Table 1 summarizes the locations and exonic structures for human *CES* genes based upon previous reports for human *CES1* and *CES2* (Becker et al. 1994; Ghosh 2000; Langmann et al. 1997; Marsh et al. 2004) and predictions for human *CES3* (Holmes et al. 2010), *CES4A* (Holmes et al. 2009a), and *CES5A* (Holmes et al. 2008a) [the February 2009 human reference sequence (GRCh37) was used in this study (Rhead et al. 2010)]. Human *CES1P1* (a *CES1*-like pseudogene), *CES1*, and *CES5A* were located in a cluster (cluster 1) on chromosome 16, while *CES2*, *CES3*, and *CES4A* were in a separate cluster (cluster 2) on the same chromosome. Cluster 1 *CES* genes (*CES1* and *CES5A*) were transcribed on the negative strand, whereas cluster 2 genes (*CES2*, *CES3*, and *CES4A*) were transcribed on the positive strand. Figure 1 summarizes the predicted exonic start sites for human *CES* genes, with *CES1* and *CES4A* containing 14 exons, *CES3* and *CES5A* 13 exons, and CES2 with 12 exons. These exon start sites were in identical or similar positions to those reported for *CES1* (Ghosh 2000; March et al. 2004). Figure 2 shows the comparative structures for human *CES* reference sequences and transcripts described on the AceView website (http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/) (Thierry-Mieg and Thierry-Mieg 2006). The *CES* gene and transcript sequences varied in size from 11 kb for *CES2* to 79 kb for *CES5A* and exhibited distinct structures in each case. Moreover, several isoforms were generated in vivo for each of the human *CES* genes and have different structures as a result of transcriptional events, including truncation of the 5' ends, differential presence or absence of exons, alternative splicing or retention of introns, or overlapping exons with different boundaries. In addition, the isoforms are differentially expressed in tissues of the body and may perform distinctive metabolic roles. *CES* isoforms were named by using the gene name followed by the GenBank ID for the specific transcript. Recent studies of human *CES1* have described at least two major isoform transcripts, designated as *CES1A1* (*AB119997*) and *CES1A2* (*AB119996*) (Tanimoto et al. 2007). These isoforms have been redesignated as *CES1_AB119997* and *CES1_AB119997*, respectively (see Table 1) and encode sequences that differ by only four amino acid residues within the N-terminal region (exon 1) (Tanimoto et al. 2007). Distinct 5'-untranslated consensus sequences for binding transcription factors were reported. They suggested differences in transcriptional regulation and functional roles in contributing to CPT-11 chemosensitivity for these isoforms (Hosokawa et al. 2008; Tanimoto et al. 2007; Yoshimura et al. 2008). Fukami et al. (2008) have also examined human *CES* isoform structure and proposed that *CES1P1*,a *CES1*-like pseudogene on chromosome 16 (designated as *CES1A3*), was derived from the *CES1_AB119997* isoform.

human CES amino acid sequences and structures

An alignment of the amino acid sequences for human CES-like protein subunits is shown in Fig. 1, together with a description of several features for these enzymes. The sequences have been derived from previously reported sequences for CES1 (Munger et al. 1991; Shibata et

al. 1993), CES2 (Langmann et al. 1997; Schewer et al. 1997), CES3 (Sanghani et al. 2004), CES4A (previously CES6 or CES8) (Holmes et al. 2009a); and CES5A (previously CES7) (Holmes et al. 2008a) (Table 1). Alignments of the human CES subunits showed between 39 and 46% sequence identities, which suggests that these are products of separate but related gene families, whereas sequence alignments of human CES1 and CES2 with mouse CES1-like and CES2-like subunits exhibited higher levels of sequence identities with the CES family homolog in each case [66–78% identities for human and mouse CES1-like subunits and 64–72% for human and mouse CES2-like subunits, respectively (data not shown)], suggesting that these are members of the same mammalian CES families, in each case. Similar results were observed for comparisons of human CES3, CES4A (previously CES6 or CES8), and CES5A (previously CES7) with the corresponding mouse CES homolog sequences, with 65, 72, and 69% identities being observed, respectively. This supports the designation of these *CES* genes as members of the same family, in each case.

The amino acid sequences for the human CES subunits examined contained 567 (CES1), 559 (CES2), 571 (CES3), 561 (CES4A), and 575 (CES5A) residues (Fig. 1). Previous studies on human CES1 have identified key residues that contribute to the catalytic, oligomeric, subcellular localization and regulatory functions for this enzyme (sequence numbers refer to human CES1). These included the catalytic triad for the active site (Ser221; Glu354; His468) (Cygler et al. 1993); disulfide bond-forming residues (Cys87/Cys116 and Cys274/Cys285) (Lockridge et al. 1987); microsomal targeting sequences, including the hydrophobic N-terminus signal peptide (Potter et al. 1998; von Heijne 1983; Zhen et al. 1995) and the C-terminal endoplasmic reticulum (ER) retention sequence (His-Ile-Glu-Leu) (Robbi and Beaufay 1983); and ligand-binding sites, including the “Z-site” (Gly356), the ”side door” (Val424-Met425-Phe426), and the “gate” (Phe550) residues (Bencharit et al. 2003, 2006; Fleming et al. 2005). Identical residues were observed for each of the human CES subunit families for the active site triad and disulfide bond-forming residues, although changes were observed for some key residues for CES1 subunits, including the “side-door” and “gate” of the active site, with family-specific sequences or residues in each case. The “Z-site” (Gly356 for human CES1) has been retained for human CES2 and CES5A sequences, but substituted for CES3 (Ser) and CES4A (Asn). The hydrophobic N-terminal sequence for human CES sequences has undergone major changes, although this region retains a predicted signal peptide property. The human CES C-terminal tetrapeptide sequences have also changed, although CES2 (HTEL) and CES3 (QEDL) are similar in sequence with human CES1 (HIEL), which plays a role in the localization of human CES1 within endoplasmic reticulum membranes (Robbi and Beaufay 1983).

Other key human CES1 sequences included two charge clamps that are responsible for subunit-subunit interaction, namely, residues Lys78/Glu183 and Glu72/Arg186, which contribute to the trimeric and hexameric structures for this enzyme (Bencharit et al. 2003, 2006; Fleming et al. 2005). Other human CES subunit sequences for these charge clamp sites included substitutions with neutral amino acids for the human CES2 and CES5A sequences, while the CES3 and CES4A sequences retained one potential clamp site (Fig. 1). Pindel et al. (1997) and Holmes et al. (2009b) have reported monomeric subunit structures for human and baboon CES2, which is consistent with the absence of charge clamps for this enzyme. This could have a major influence on the kinetics and biochemical roles for human CES isozymes since three-dimensional studies have indicated that ligand binding to the human CES1 “Z-site” shifts the trimer-hexamer equilibrium toward the trimer that facilitates substrate binding and enzyme catalysis (Redinbo and Potter 2005). The *N*-glycosylation site for human CES1 at Asn79-Ala80-Thr81 (Bencharit et al. 2003, 2006; Fleming et al. 2005; Kroetz et al. 1993) was not retained for any of the other human CES sequences, although potential *N*-glycosylation sites were observed at other positions, including CES2 (site 3), CES3 (site 2), CES4A (sites 4, 5, and 7), and CES5A (sites 6, 8, and 9) (Table 4). Given the

reported role of the *N*-glycosylated carbohydrate group contributing to CES1 stability and maintaining catalytic efficiency (Kroetz et al. 1993), the *N*-glycosylation sites predicted for other human CES subunits may perform similar functions or indeed may serve new functions specific to a particular CES family.

Predicted secondary structures for human CES2 (Holmes et al. 2009b), CES3 (Holmes et al. 2010), CES4A (Holmes et al. 2009a), and CES5A (Holmes et al. 2008a) sequences were compared with those reported for human CES1, and similar α-helix β-sheet structures were observed for all of the CES subunits examined (Bencharit et al. 2003, 2006) (Fig. 1). This was especially apparent near key residues or functional domains such as the α-helix within the N-terminal signal peptide, the β-sheet and α-helix structures near the active site Ser221 (human CES1) and "Z-site" (Glu354/Gly356, respectively), the α-helices bordering the "side door" site, and the a-helix containing the "gate" residue (Phe550 for human CES1). The human CES5A sequence, however, contained a predicted helix at the hydrophobic C-terminus not observed for other CES subunits which may perform a family-specific function. Predicted 3D structures have been previously described for each of the human CES subunits (Holmes et al. 2008a, 2009a, b, 2010); they were similar to the human CES1 structure (Bencharit et al. 2003, 2006).

Mouse *Ces* genes and enzymes

Table 2 summarizes the proposed names, locations, and overall structures for the *Ces* genes observed for the mouse genome (July 2007 mouse [*Mus musculus*] genome data obtained from the Build 37 assembly by NCBI and the Mouse Genome Sequencing Consortium) (http://www.ncbi.nlm.nih.gov was used in this study). The italicized gene name *Ces* is consistent with other mouse gene nomenclature and is preferred to the *CES* stem used for human genes. At least 20 mouse *Ces* genes are recognized on the Mouse Genome Database http://www.informatics.jax.org/) (MGI) and further described in terms of their locations on mouse chromosome 8, the number of predicted exons for each gene, predicted strand for transcription, number of amino acid residues and subunit molecular weights (MWs) for the encoded CES subunits, and identification symbols from MGI (e.g., MGI3648919 for *Ces1a*), NCBI (Reference Sequences were identified from the National Center for Biotechnology Information database) (http://www.ncbi.nlm.nih.gov/), Vega (the VErtebrate Genome Annotation database) (http://vega.sanger.ac.uk/index.html), UNIPROT (Universal Protein Resource) (http://www.ebi.ac.uk/uniprot/), and Ensembl (Genome Database) (http://www.ensembl.org/) database sources.

Eight *Ces1*-like genes are located in tandem within a 360-kb segment of mouse chromosome 8, with an average gene size of 28 kb. The names for these genes (*Ces1a*, *Ces1b*,…, *Ces1h*) are allocated in the same order as their locations on the mouse genome (Table 3). The *Ces1*-like gene cluster is also located near the mouse *Ces5a* gene, which is comparable to the *CES1P1-CES1-CES5A* cluster observed for human chromosome 16. Each of these genes contained 13 or 14 exons predicted for transcription on the negative strand and with encoded CES subunits exhibiting distinct but similar amino acid sequences (554–567 residues). The subunits were 63–85% identical with each other and with the human CES1 sequence, which is consistent with these being members of the mouse *Ces1* gene family. Mouse *Ces1*-like genes included several that have been previously investigated, including *Ces1c* (previously called Es1), encoding a major mouse plasma esterase with 554 amino acid residues and also exhibiting lung surfactant convertase activity (Genetta et al. 1988; Krishnasamy et al. 1998); *Ces1d* (previously *Ces3*), encoding a mouse liver enzyme with 565 residues and exhibiting triacylglycerol hydrolase activity (Dolinsky et al. 2001); *Ces1e* (previously called *Es*22 or egasyn), encoding a liver CES with 562 residues and exhibiting β-glucuronidase-binding properties (Ovnic et al. 1991); and *Ces1g* (previously *Ces1*), encoding a liver CES with 565

amino acid residues and exhibiting lipid metabolizing activity (Table 4) (Ellingham et al. 1998).

Eight *Ces2*-like genes were also observed in a second 286-kb gene cluster on mouse chromosome 8, with an average gene size of approximately 8 kb (Table 2). These genes were named according to their sequence of position on the mouse genome (*Ces2a*, *Ces2b*, …, *Ces2h*) and included a pseudogene designated *Ces2d-ps*. Three of these mouse *Ces2*-like genes have been previously described, including *Ces2c* (previously *Ces2*), encoding an inducible liver acyl-carnitine hydrolase enzyme with 561 residues (Furihata et al. 2003); *Ces2e* (previously *Ces5*), encoding a liver and intestinal enzyme with 560 amino acid residues (The MGC Project Team 2004); and *Ces2a* (previously *Ces6*), encoding a liver and colon enzyme with 558 residues (The MGC Project Team 2004). The *Ces2*-like cluster was located alongside two *Ces3*-like mouse genes (*Ces3a* and *Ces3b*) and a *Ces4a* gene (Table 3); this is comparable to the *CES2-CES3-CES4A* gene cluster on human chromosome 16 (Table 1). The *Ces3a* gene (previously mouse *esterase 31* or *Est31*) is expressed strongly in male mouse livers and encodes a 554-residue CES3-like subunit (Aida et al. 1993), whereas the *Ces3b* gene (previously *Es31L* or *EG13909*) is also expressed in liver and encodes a 568-residue subunit (The MGC Project Team 2004). The *Ces4a* gene (previously called *EST8* or *Ces8*) encodes an enzyme predicted for secretion in epidermal cells with 563 amino acid residues and showing 72% identity with human CES4A (The MGC Project Team 2004).

Rat *Ces* genes and enzymes

Table 3 summarizes the proposed names, locations, and structures for *Ces* genes observed for the rat genome [the November 2004 rat (*Rattus norvegicus*) genome assembly based on version 3.4 produced by the Baylor Human Genome Sequencing Center (Gibbs et al. 2004) was used in this study]. Fifteen rat *Ces* genes were identified on the Rat Genome Database (RGD) (http://rgd.mcw.edu/) and further characterized by their locations on rat chromosomes 1 and 19, the number of predicted exons for each gene, the predicted strand for transcription, current gene symbols, the number of amino acid residues and subunit MWs for the encoded CES subunits, and the identification symbols from RGD (e.g., RGD1583671 for *Ces1a*), NCBI Reference Sequences (http://www.ncbi.nlm.nih.gov/), Vega (http://vega.sanger.ac.uk/index.html), UNIPROT (http://www.ebi.ac.uk/uniprot/), and Ensembl (http://www.ensembl.org/) database sources.

Five *Ces1*-like genes were located in tandem within a 201-kb segment of rat chromosome 19, with an average gene size of 33 kb (Table 3). The names for these genes (*Ces1a*, *Ces1c*, …, *Ces1f*) were allocated according to their degree of identity with the corresponding mouse *Ces1*-like genes (Table 3). The genes were located in tandem in the same order as the mouse *Ces1*-like genes and were near the rat *Ces5a* gene. This is comparable to the *CES1P1–CES1A–CES5A* gene cluster observed for human chromosome 16. The rat *Ces1*-like genes contained 14 exons and were predicted for transcription on the positive strand, with encoded CES subunits exhibiting similar amino acid sequences (550–565 residues). The subunits were 65–73% identical with each other and with the human CES1 sequence, which is consistent with membership of the rat *Ces1* gene family. The encoded rat *Ces1*-like subunit sequences showed higher levels of identity with the corresponding mouse *Ces1*-like sequences (81–92% for rat and mouse CES1a, CES1c, CES1d, CES1e, and CES1f amino acid sequences). At least three rat *Ces1*-like genes have been previously described, including *Ces1c* (previously called *Es1*), encoding a rat plasma esterase (Sanghani et al. 2002; Vanlith et al. 1993); *Ces1d* (previously *Ces3*), encoding a rat liver enzyme with 565 residues and exhibiting cholesteryl ester hydrolase activity (Ghosh et al. 1995; Robbi et al. 1990); and

Ces1e (previously called *ES-3* or *egasyn*), encoding a rat liver Ces with 561 residues and having β-glucuronidase-binding properties (Robbi and Beaufay 1994).

Seven rat *Ces2*-like genes were observed on the rat genome and were localized on two chromosomes: chromosome 1 (*Ces2c* and *Ces2i*) and chromosome 19 in three locations: *Ces2a* and *Ces2e*; *Ces2j*; and *Ces2g* and *Ces2h* (Table 3). The genes were named according to the degree of sequence identity with the corresponding mouse *Ces2*-like genes. Rat *Ces2*-like genes have been previously investigated, including *Ces2c* (previously *Ces2*), encoding an inducible liver acyl-carnitine hydrolase enzyme with 561 residues (Furihata et al. 2003); *Ces2e* (previously *Ces5*), encoding a liver and intestinal enzyme with 560 amino acid residues (The MGC Project Team 2004); and *Ces2a* (previously *Ces6*), encoding a liver and colon enzyme with 558 residues.(The MGC Project Team 2004). The rat *Ces2*-like cluster was located alongside a *Ces3*-like gene (*Ces3a* and *Ces3b*) and a *Ces4a* gene (Table 3), which is comparable to the *CES2A-CES3A-CES4A* gene cluster on human chromosome 16 (Table 1).

Functions of mammalian *CES* families

Mammalian CES families exhibit broad substrate specificities, and specific roles for these enzymes have been difficult to establish because of the promiscuity of the CES active site toward a wide range of substrates and the existence of multiple forms with overlapping specificities (Fleming et al. 2005; Imai 2006; Leinweber 1987; Redinbo and Potter 2005; Satoh and Hosokawa 1998, 2006). Table 4 summarizes current knowledge concerning substrates and functions reported for human, mouse, and rat *CES* gene family members.

Studies on human CES1 have examined its role in the metabolism of various drugs, including narcotics such as heroin and cocaine (Bencharit et al. 2003; Pindel et al. 1997), warfare nerve agents (Hemmert et al. 2010), psy-chostimulants (Sun et al. 2004), analgesics (Takai et al. 1997), and chemotherapy drugs (Sanghani et al. 2004). Mammalian liver is predominantly responsible for drug clearance from the body, with CES1 and CES2 (with CES1 > CES2) playing major roles, following absorption of drugs into the circulation (Imai 2006; Pindel et al. 1997). Mammalian intestine (with CES2 > CES1) plays a major role in first-pass clearance of several drugs, predominantly via CES2 in the ileum and jejunum (Imai et al. 2003). CES1 and CES2 also have different roles in prodrug activation, as shown for the anticancer drug irinotecan (CPT-11), which is converted to its active form SN-38 predominantly by CES2 (Humerickhouse et al. 2000). Recent modeling studies have shown that the human CES2 active site cavity is lined with negatively charged residues; this may explain the preference of this enzyme for neutral substrates (Vistoli et al. 2010). The role for human CES3 has not been studied extensively, although the enzyme is capable of activating prodrugs such as irinotecan (Sanghani et al. 2004). There are no reports concerning the metabolic role(s) for human CES4A, and functional studies on mammalian CES5 function are limited to feline species, where the enzyme is secreted into cat urine and apparently regulates the production of a cat-specific amino acid “felinine,” a putative pheromone precursor (Miyazaki et al. 2006).

Evolution of mammalian *CES* gene families

Recent comparative and evolutionary studies (Holmes et al. 2008b; Williams et al. 2010) have concluded that there are at least five major mammalian *CES* gene families. In addition, the gene duplication events that generated the ancestral mammalian *CES1*, *CES2*, *CES3*, *CES4*, and *CES5* genes have apparently predated the common ancestor for marsupial and eutherian mammals (Holmes et al. 2008b) which has been estimated at approximately 173–193 million years ago (Woodburne et al. 2003) and may coincide with the early diversification of tetrapods approximately 350–360 million years ago (Donoghue and

Benton 2007). The mammalian *CES* gene families are ancient in their genetic origins and were established prior to the appearance of mammals during evolution. Further *CES/Ces* gene duplication events have subsequently occurred during mammalian evolution, however, especially for rodent species, for which the mouse and rat *Ces1*-like and *Ces2*-like genes have apparently undergone successive duplication events. At least three of these are likely to have occurred in the common ancestor for rat and mouse during rodent evolution since several homolog genes and proteins were recognized, including *Ces1c* (previously Es1), *Ces1d* (*Ces3*), *Ces1e* (*Es22*), *Ces2a* (*Ces6*), *Ces2c* (*Ces2*), and *Ces2e* (*Ces5*) (Tables 3, 4). With the exception of the rat *Ces2*-like genes, which were located in multiple clusters on chromosomes 1 and 19, human, mouse, and rat *CES* genes were localized within two clusters of genes on the same chromosome, namely, *Ces1–Ces5A* (with multiple *Ces1*-like genes) and *Ces2–Ces3–Ces4A* (with multiple *Ces2*-like genes in mouse and rat). The presence of two *Ces3*-like genes in the mouse suggests that a further duplication event also took place in this species.

Conclusions

This article has examined human, mouse, and rat carboxylesterase genes and encoded subunits and has proposed a new nomenclature system, identifying each of five gene families (designated as *CES1*, *CES2*,…, *Ces5* for human genes and *Ces1*, *Ces2*,…, *Ces5* for mouse and rat genes) and allocating a unique gene name for each of the genes. The italicized root symbol "*CES*" for human and "*Ces*" for mouse and rat genes followed by a number for the family were used, which is consistent with current practice. When multiple genes were identified for a gene family or where a gene required a name that clashed with an existing name, a capital letter (for human genes) (e.g., *CES4A*) or a lower-case letter (for mouse and rat genes) (e.g., *Ces1a*, *Ces1b*) was added after the number. A human *CES* pseudogene was named, using a capital "P" and a number (e.g., *CES1P1*), whereas mouse and rat *Ces* pseudogenes were named with a unique lower-case letter followed by "-ps" (e.g., *Ces2d-ps*). This new nomenclature will also assist in naming multiple CES genes and proteins from other mammalian species. As an example, Holmes et al. (2009c) and Williams et al. (2010) have reported multiple *CES1*-like genes on the horse genome that may be designated in accordance with the recommended nomenclature as *CES1A*, *CES1B*, *CES1C*, and so on, in order of the tandem locations of these genes on chromosome 3. Transcript isoforms of *CES* gene transcripts were named by following the gene name with the GenBank ID for the specific transcript. This nomenclature will assist our understanding of the genetic relatedness and the CES family origins for individual human, mouse, and rat *CES* genes and proteins and facilitate future research into the structure, function, and evolution of these genes. It will also serve as a model for naming *CES* genes from other mammalian species.

Acknowledgments

This research was supported by NIH Grants P01 HL028972 and P51 RR013986 (to LAC); R01 ES07965 (to BY); and CA108775, and a Cancer Center Core Grant CA21765, the American Lebanese and Syrian Associated Charities (ALSAC) and St. Jude Children's Research Hospital (SJCRH) (to PMP); and a program project grant HG000330 entitled `Mouse Genome Informatics' from the National Human Genome Research Institute of the NIH (to LJM). Acknowledgement is also given to members of the Redinbo laboratory and NIH grants CA98468 and NS58089 (to MRR).

References

Aida K, Moore R, Negishi M. Cloning and nucleotide sequence of a novel, male-predominant carboxylesterase in mouse liver. Biochim Biophys Acta. 1993; 1174:72–74. [PubMed: 7916639]

Barthel BL, Torres RC, Hyatt JL, Edwards CC, Hatfield MJ, et al. Identification of human intestinal carboxylesterase as the primary enzyme for activation of a doxazoline carbamate prodrug. J Med Chem. 2008; 51:298–304. [PubMed: 18173233]

Becker A, Bottcher A, Lackner KJ, Fehringer P, Notka F, et al. Purification, cloning and expression of a human enzyme with acyl coenzyme A: cholesterol acyltransferase activity, which is identical to liver carboxylesterase. Arterioscler Thromb. 1994; 14:1346–1355. [PubMed: 8049197]

Bencharit S, Morton CL, Xue Y, Potter PM, Redinbo MR. Structural basis of heroin and cocaine metabolism by a promiscuous human drug-processing enzyme. Nat Struct Biol. 2003; 10:349–356. [PubMed: 12679808]

Bencharit S, Edwards CC, Morton CL, Howard-Williams EL, Kuhn P, et al. Multisite promiscuity in the processing of endogenous substrates by human carboxylesterase 1. J Mol Biol. 2006; 363:201–214. [PubMed: 16962139]

Berning W, De Looze SM, von Deimling O. Identification and development of a genetically closely linked carboxylesterase gene family of the mouse liver. Comp Biochem Physiol. 1985; 80:859–865.

Cygler M, Schrag JD, Sussman JL, Harel M, Silman I, et al. Relationship between sequence conservation and three-dimensional structure in a large family of esterases, lipases and related proteins. Protein Sci. 1993; 2:366–382. [PubMed: 8453375]

Diczfalusy MA, Bjorkkem I, Einarsson C, Hillebrant CG, Alexson SE. Characterization of enzymes involved in formation of ethyl esters of long-chain fatty acids. J Lipid Res. 2001; 42:1025–1032. [PubMed: 11441128]

Dolinsky VW, Sipione S, Lehner R, Vance DE. The cloning and expression of murine triacylglycerol hydrolase cDNA and the structure of the corresponding gene. Biochim Biophys Acta. 2001; 1532:162–172. [PubMed: 11470237]

Donoghue PCJ, Benton MJ. Rocks and clocks: calibrating the tree of life using fossils and molecules. Trends Genet. 2007; 22:424–630.

Ecroyd H, Belghazi M, Dacheux JL, Miyazaki M, Yamashita T, et al. An epididymal form of cauxin, a carboxylesterase-like enzyme, is present and active in mammalian male reproductive fluids. Biol Reprod. 2006; 74:439–447. [PubMed: 16251497]

Ellingham P, Seedorf U, Assmann G. Cloning and sequencing of a novel murine liver carboxylesterase cDNA. Biochim Biophys Acta. 1998; 1397:175–179. [PubMed: 9565681]

Fleming CD, Bencharit S, Edwards CC, Hyatt JL, Tsurkan L, et al. Structural insights into drug processing by human carboxylesterase 1: tamoxifen, Mevastatin, and inhibition by Benzil. J Mol Biol. 2005; 352:165–177. [PubMed: 16081098]

Fukami T, Nakajima M, Maruichi T, Takahashi S, Takamiya M, et al. Structure and characterization of human carboxylesterase 1A1, 1A2 and 1A3 genes. Pharm Genomics. 2008; 18:911–920.

Furihata T, Hosokawa M, Nakata F, Satoh T, Chiba K. Purification, molecular cloning, and functional expression of inducible liver acylcarnitine hydrolase in C57BL/6 mouse, belonging to the carboxylesterase multigene family. Arch Biochem Biophys. 2003; 416:101–109. [PubMed: 12859986]

Genetta TL, D'Eustachio P, Kadner SS, Finlay TH. cDNA cloning of esterase 1, the major esterase activity in mouse plasma. Biochem Biophys Res Commun. 1988; 151:1364–1370. [PubMed: 2895647]

Ghosh S. Cholesteryl ester hydrolase in human monocyte/macrophage: cloning, sequencing and expression of full-length cDNA. Physiol Genomics. 2000; 2:1–8. [PubMed: 11015575]

Ghosh S, Mallonee DH, Grogan WM. Molecular cloning and expression of rat hepatic neutral cholesteryl ester hydrolase. Biochim Biophys Acta. 1995; 1259:305–312. [PubMed: 8541339]

Gibbs RA, Weinstock GM, Metzker ML, Muzny DM, Sodergren EJ, et al. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. Nature. 2004; 428:493–521. [PubMed: 15057822]

Gilham D, Alam M, Gao W, Vance DE, Lehner R. Triacylglycerol hydrolase is localized to the endoplasmic reticulum by an unusual retrieval sequence where it participates in VLDL assembly without utilizing VLDL lipids as substrates. Mol Biol Cell. 2005; 16:984–996. [PubMed: 15601899]

Hemmert AC, Otto TC, Wierdl M, Edwards CC, Fleming CD, et al. Human carboxylesterase 1 stereoselectively binds the nerve agent cyclosarin and spontaneously hydrolyzes the nerve agent sarin. Mol Pharmacol. 2010; 77:508–516. [PubMed: 20051531]

Holmes RS, Cox LA, VandeBerg JL. Mammalian carboxylesterase 5: comparative biochemistry and genomics. Comp Biochem Physiol D Genomics Proteomics. 2008a; 3:195–204.

Holmes RS, Chan J, Cox LA, Murphy WJ, VandeBerg JL. Opossum carboxylesterases: sequences, phylogeny and evidence for CES duplication events predating the marsupial-eutherian common ancestor. BMC Evol Biol. 2008b; 8:54. [PubMed: 18289373]

Holmes RS, VandeBerg JL, Cox LA. A new class of mammalian carboxylesterase *CES6*. Comp Biochem Physiol Part D Genomics Proteomics. 2009a; 4:209–217. [PubMed: 20161041]

Holmes RS, Glenn JP, VandeBerg JL, Cox LA. Baboon carboxylesterases 1 and 2: sequences, structures and phylogenetic relationships with human and other primate carboxylesterases. J Med Primatol. 2009b; 38:27–38. [PubMed: 19187434]

Holmes RS, Cox LA, VandeBerg JL. Horse carboxylesterases: evidence for six CES1 and four families of CES genes on chromosome 3. Comp Biochem Physiol. 2009c; 4:54–65.

Holmes RS, Cox LA, VandeBerg JL. Mammalian carboxylesterase 3: comparative genomics and proteomics. Genetica. 2010; 138(7):695–708. [PubMed: 20422440]

Hosokawa M. Structure and catalytic properties of carboxylesterase isozymes involved in metabolic activation of prodrugs. Molecules. 2008; 13:412–431. [PubMed: 18305428]

Hosokawa M, Furihata T, Yaginuma Y, Yamamoto N, Kayano N, et al. Genomic structure and transcriptional regulation of the rat, mouse and human carboxylesterase genes. Drug Metab Rev. 2007; 39:1–15. [PubMed: 17364878]

Hosokawa M, Furihata T, Yaginuma Y, Yamamoto N, Watanabe N, et al. Structural organization and characterization of the regulatory element of the human carboxylesterase (*CES1A1* and *CES1A2*) genes. Drug Metab Pharmacokinet. 2008; 23:73–84. [PubMed: 18305377]

Humerickhouse R, Lohrbach K, Li L, Bosron WF, Dolan ME. Characterization of CPT-11 hydrolysis by human liver carboxylesterase isoforms h-CE1 and hCE-2. Cancer Res. 2000; 60:1189–1192. [PubMed: 10728672]

Imai T. Human carboxylesterase isozymes: catalytic properties and rational drug design. Drug Metab Pharmacokinet. 2006; 21:173–185. [PubMed: 16858120]

Imai T, Yoshigae Y, Hosokawa M, Chiba K, Otagiri M. Evidence for the involvement of a pulmonary first-pass effect via carboxylesterase in the disposition of a propanolol ester derivative after intravenous administration. J Pharmacol Exp Ther. 2003; 307:1234–1242. [PubMed: 14534358]

Ko KW, Erickson B, Lehner R. Es-x/Ces1 prevents triacylglycerol accumulation in McArdle-RH7777 hepatocytes. Biochim Biophys Acta. 2009; 1791:1133–1143. [PubMed: 19651238]

Krishnasamy R, Teng AL, Dhand R, Schultz RM, Gross NJ. Molecular cloning, characterization and differential expression pattern of mouse lung surfactant convertase. Am J Physiol Lung Mol Cell Biol. 1998; 275:L969–L975.

Kroetz DL, McBride OW, Gonzalez FJ. Glycosylation-dependent activity of Baculovirus-expressed human liver carboxylesterases: cDNA cloning and characterization of two highly similar enzyme forms. Biochemistry. 1993; 32:11606–11617. [PubMed: 8218228]

Langmann T, Becker A, Aslanidis C, Notka F, Ulrich H, et al. Structural organization and characterization of the promoter region of a human carboxylesterase gene. Biochim Biophys Acta. 1997; 1350:65–74. [PubMed: 9003459]

Lehner R, Vance DE. Cloning and expression of a cDNA encoding a hepatic microsomal lipase that mobilizes stored triacylglycerol. Biochem J. 1999; 343:1–10. [PubMed: 10493905]

Leinweber FJ. Possible physiological roles of carboxyl ester hydrolases. Drug Metab Rev. 1987; 18:379–439. [PubMed: 3286170]

Linke T, Dawson H, Harrison EH. Isolation and characterization of a microsomal retinyl ester hydrolase. J Biol Chem. 2005; 280:23287–23294. [PubMed: 15767260]

Lockridge O, Adkins S, La Due BN. Location of disulfide bonds within the sequence of human serum cholinesterase. J Biol Chem. 1987; 262:12945–12952. [PubMed: 3115973]

Marsh S, Xiao M, Yu J, Ahluwalia R, Minton M, et al. Pharmacogenomic assessment of carboxylesterases 1 and 2. Genomics. 2004; 84:661–668. [PubMed: 15475243]

Masaki K, Hashimoto M, Imai T. Intestinal first-pass metabolism via carboxylesterase in rat jejunum and intestine. Drug Metab Dispos. 2007; 35:1089–1095. [PubMed: 17392394]

Miyazaki M, Kamiie K, Soeta S, Taira H, Yamashita T. Molecular cloning and characterization of a novel carboxylest-erase-like protein that is physiologically present at high concentrations in the urine of domestic cats (*Felis catus*). Biochem J. 2003; 370:101–110. [PubMed: 12401131]

Miyazaki M, Yamashita T, Suzuki Y, Saito Y, Soeta S, et al. A major urinary protein of the domestic cat regulates the production of felinine, a putative pheromone precursor. Chem Biol. 2006; 13:1070–1079.

Morton CL, Iacono L, Hyatt JL, Taylor KR, Cheshire PJ, et al. Activation and antitumor activity of CPT-11 in plasma esterasedeficient mice. Cancer Chemother Pharmacol. 2005; 56:629–636. [PubMed: 15918039]

Munger JS, Shi GP, Mark EA, Chin DT, Gerard C, et al. A serine esterase released by human alveolar macrophages is closely related to liver microsomal carboxylesterases. J Biol Chem. 1991; 266:18832–18838. [PubMed: 1918003]

Mutch E, Nave R, McCracken N, Zech K, Williams FM. The role of esterases in the metabolism of ciclesinide to deisobutyrlciclesonide in human tissue. Biochem Pharmacol. 2007; 73:1657–1664. [PubMed: 17331475]

Ohtsuka H, Inoue S, Kameyama M. Intracellular conversion of irinotecan to its active form, SN-38, by native carboxylesterase in human non-small cell lung cancer. Lung Cancer. 2003; 41:87–198.

Okazaki H, Igarashi M, Nishi M, Tajima M, Sekiya M, et al. Identification of a novel member of the carboxylesterase family that hydrolyzes triacylglycerol. A potential role in adipocyte lipolysis. Diabetes. 2006; 55:2091–2097. [PubMed: 16804080]

Okazaki H, Igarashi M, Nishi M, Sekiya M, Tajima M, et al. Identification of neutral cholesterol hydrolase, a key enzyme removing cholesterol from macrophages. J Biol Chem. 2008; 283:33357–33364. [PubMed: 18782767]

Ovnic M, Swank RT, Fletcher C, Zhen L, Novak EK, et al. Characterization and functional expression of a cDNA encoding egasyn (esterase-22): the endoplasmic reticulum-targeting protein of beta-glucuronidase. Genomics. 1991; 11:956–967. [PubMed: 1783403]

Pindel EV, Kedishvili NY, Abraham TL, Brezinski MR, Zhang A, et al. Purification and cloning of a broad substrate specificity human liver carboxylesterase that catalyzes the hydrolysis of cocaine and heroin. J Biol Chem. 1997; 272:14769–14775. [PubMed: 9169443]

Potter PM, Wolverton JS, Morton CL, Wierdl M, Danks MK. Cellular localization domains of a rabbit and human carboxylesterase: influence on irinotecan (CPT-11) metabolism by the rabbit enzyme. Cancer Res. 1998; 58:3627–3632. [PubMed: 9721871]

Redinbo MR, Potter PM. Mammalian carboxylesterases: from drug targets to protein therapeutics. Drug Discov Today. 2005; 10:313–320. [PubMed: 15749280]

Rhead B, Karolchik D, Kuhn RM, Hinrichs AS, Zweig AS, et al. The UCSC Genome Browser database: update 2010. Nucl Acids Res. 2010; 38:D613–D619. [PubMed: 19906737]

Robbi M, Beaufay H. Purification and characterization of various esterases from rat liver. Eur J Biochem. 1983; 137:293–301. [PubMed: 6653557]

Robbi M, Beaufay H. Cloning and sequencing of rat liver carboxylesterase ES-3 (egasyn). Biochem Biophys Res Commun. 1994; 203:1404–1411. [PubMed: 7945287]

Robbi M, Beaufay H, Octave JN. Nucleotide sequence of cDNA coding for rat liver pI 6.1 esterase (ES-10), a carboxylesterase located in the lumen of the endoplasmic reticulum. Biochem J. 1990; 269:451–458. [PubMed: 2386485]

Ruppert C, Bagheri A, Markart P, Schmidt R, Seegar W, et al. Liver carboxylesterase cleaves surfactant protein (SP-B) and promotes surfactant subtype conversion. Biochem Biophys Res Commun. 2006; 348:1449–1454. [PubMed: 16919595]

Sanghani SP, Davis WI, Dumaual NG, Mahrenholz A, Bosron WF. Identification of microsomal rat liver carboxylesterases and their activity with retinyl palmitate. Eur J Biochem. 2002; 269:4387–4398. [PubMed: 12230550]

Sanghani SP, Quinney SK, Fredenberg TB, Davis WI, Murray DJ, et al. Hydrolysis of irinotecan and its oxidative metabolites, 7-ethyl-10-[4-N(5-aminopentanoic acid)-1-piperidino] carbonyloxycampothecin and 7-ethyl-10-[4-(1-piperidino)-1 amino]-carbonyloxycamptothecin, by

human carboxylesterases CES1A1, CES2, and a newly expressed carboxylesterase isoenzyme, CES3. Drug Metab Dispos. 2004; 32:505–511. [PubMed: 15100172]

Satoh T, Hosokawa M. Molecular aspects of carboxylesterase isoforms in comparison with other esterases. Toxicol Letters. 1995; 82–83:439–445.

Satoh T, Hosokawa M. The mammalian carboxylesterases: from molecules to functions. Ann Rev Pharmacol Toxicol. 1998; 38:257–288. [PubMed: 9597156]

Satoh T, Hosokawa M. Structure, function and regulation of carboxylesterases. Chem Biol Interact. 2006; 162:195–211. [PubMed: 16919614]

Satoh T, Taylor P, Bosron WF, Sanghani P, Hosokawa M, et al. Current progress on esterases: from molecular structure to function. Drug Metab Dispos. 2002; 30:488–493. [PubMed: 11950776]

Schewer H, Langmann T, Daig R, Becker A, Aslandis C, et al. Molecular cloning and characterization of a novel putative carboxylesterase, present in human intestine and liver. Biochem Biophys Res Commun. 1997; 233:117–120. [PubMed: 9144407]

Schreiber R, Taschler U, Wolinski H, Seper A, Tamegger SN, et al. Esterase 22 and beta-glucuronidase hydrolyze retinoids in mouse liver. J Lipid Res. 2009; 50:2514–2523. [PubMed: 19723663]

Shibita F, Takagi Y, Kitajima M, Kuroda T, Omura T. Molecular cloning and characterization of a human carboxylesterase gene. Genomics. 1993; 17:76–82. [PubMed: 8406473]

Sun Z, Murry DJ, Sanghani SP, Davis WI, Kedishvilli NY, et al. Methylphenadate is stereoselectively hydrolyzed by human carboxylesterase CES1A1. J Pharmcol Exp Ther. 2004; 310:469–476.

Takai S, Matsuda A, Usami Y, Adachi T, Sugiyama T, et al. Hydrolytic profile for ester- or amide-linkage by carboxylesterases pI 5.3 and 4.5 from human liver. Biol Pharm Bull. 1997; 20:869–873. [PubMed: 9300133]

Taketani M, Shii M, Ohura K, Ninomiya S, Imai T. Carboxylesterase in the liver and small intestine of experimental animals and human. Life Sci. 2007; 81:924–932. [PubMed: 17764701]

Tanimoto K, Kaneyasu M, Shimokuni T, Hiyama K, Nishiyama M. Human carboxylesterase 1A2 expressed from carboxylesterase 1A1 and 1A2 genes is a potent predictor of CPT-11 cytotoxicity in vitro. Pharm Genomics. 2007; 17:1–10.

The MGC Project Team. The status, quality, and expansion of the NIH full-length cDNA project: the Mammalian Gene Collection (MGC). Genome Res. 2004; 14:2121–2127. [PubMed: 15489334]

Thierry-Mieg D, Thierry-Mieg J. AceView: a comprehensive cDNA-supported gene and transcripts annotation. Genome Biol. 2006; 7(Suppl 1):S12–S14. [PubMed: 16925834]

Tsujita T, Okuda H. Palmitoyl-coenzyme A hydrolyzing activity in rat kidney and its relationship with carboxylesterase. J Lipid Res. 1993; 34:1773–1781. [PubMed: 7902406]

Vanlith HA, Haller M, Vanhoof IJM, Vanderwouw MJA, Vanzutphen BFM, et al. Characterization of rat plasma esterase ES-1A concerning its molecular and catalytic properties. Arch Biochem Biophys. 1993; 301:265–274. [PubMed: 8460939]

Vistoli G, Pedretti A, Mazzolari A, Testa B. Homology modelling and metabolism prediction of human carboxylesterase-2 using docking analyses by GriDock: a parallelized tool based on AutoDock 4.0. J Comput Aided Mol Des. 2010; 24(9):771–787. [PubMed: 20623318]

von Heijne G. Patterns of amino acids near signal-sequence cleavage sites. Eur J Biochem. 1983; 133:17–21. [PubMed: 6852022]

Wang H, Gilham D, Lehner R. Proteomic and lipid characterization of apo-lipoprotein B-free luminal lipid droplets from mouse liver microsomes: implications for very low density lipoprotein assembly. J Biol Chem. 2007; 282:33218–33226. [PubMed: 17848546]

Williams ET, Wang H, Wrighton SA, Qian YW, Perkins EJ. Genomic analysis of the carboxylesterases: identification and classification of novel forms. Mol Phylogenet Evol. 2010; 57(1):23–34. [PubMed: 20510380]

Woodburne MO, Rich TH, Springer MS. The evolution of tribospheny and the antiquity of mammalian clades. Mol Phylogenet Evol. 2003; 28:360–385. [PubMed: 12878472]

Xu G, Zhang W, Ma MK, MacLeod HL. Human carboxylesterase 2 is commonly expressed in tumor tissue and is correlated with the activation of irinotecan. Clin Cancer Res. 2002; 8:2605–2611. [PubMed: 12171891]

Yan B, Matoney L, Yang D. Human carboxylesterases in term placenta: enzymatic characterization, molecular cloning and evidence for the existence of multiple forms. Placenta. 1999; 20:517–525.

Yoshimura M, Kimura T, Ishii M, Ishii K, Matsuura T, et al. Functional polymorphisms in carboxylesterase1A2 (*CES1A2*) gene involves specific protein 1 (Sp1) binding sites. Biochem Biophys Res Commun. 2008; 369:939–942. [PubMed: 18328811]

Zhang L, Hu Z, Zhu C, Liu Q, Zhou Y, et al. Identification and characterization of an epididymis-specific gene, *Ces7*. Acta Biochim Biophys Sin. 2009; 41:809–815. [PubMed: 19779645]

Zhen L, Rusiniak ME, Swank RT. The beta-glucuronidase propeptide contains a serpin-related octamer necessary for complex formation with egasyn esterase and for retention within the endoplasmic reticulum. J Biol Chem. 1995; 270:11912–11920. [PubMed: 7744842]

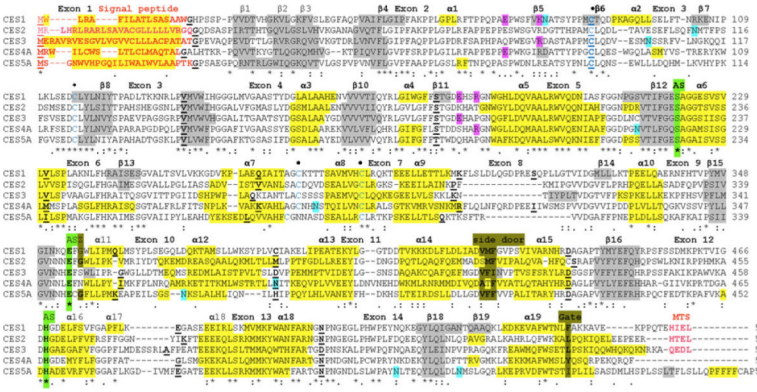

Fig. 1.

Amino acid sequence alignments for human *CES1*, *CES2*, *CES3*, *CES4A*, and *CES5A* subunits. See Table 1 for CES isoform sequences aligned. *Asterisk* identical residues for CES subunits; *colon* similar alternate residues; *dot* dissimilar alternate residues. Signal peptide sequences for CES1 (1–17), CES2 (1–25), CES3 (1–27), CES4A (1–19), and CES5A (1–24) and C- termini (MTS) microsomal targeting sequences for CES1 (564–567), CES2 (556–569), and CES3 (568–571) are shown in *red*. Active site (AS) triad residues (human CES1) Ser221, Glu354, and His468 are highlighted in *green*. “Side door” (Val424-Met425-Phe426), “Gate” (Phe550), and cholesterol binding residue (“Z site”) (Gly356) for human CES1 (Fleming et al. 2005) are highlighted in *khaki*. Disulfide bond Cys residues for human CES1 (*filled circle*) are shown in *blue*. Charge clamp residues identified for human CES1 (Glu72…Arg186; Lys78…Glu183) (Fleming et al. 2005) are highlighted in *purple*. Confirmed (CES1) (Asn79-Ala80-Thr81) [site 1] or predicted N-glycosylation sites for human CES2 (Asn111-Met112-Thr113) [site 3]; CES3 (Asn105-Ser106-Ser107) [site 2]; CES4A (Asn213-Val214-Thr215) [site 4], Asn276-Ser-277-Thr278) [site 5], and Asn388-Ile389-Thr390) [site 7]; and CES5A (Asn363-Lys364-Ser365) [site 6], (Asn513-Leu514-Thr515) [site 8], and (Asn524-Met525-Ser526 [site 9] are highlighted in blue. α-Helix (human CES1 or predicted) and β-sheet (human CES1 or predicted) regions are highlighted in *yellow* and *gray*, respectively. α-Helices and β-sheets are numbered according to the reported human CES1 3D structure (Fleming et al. 2005). *Bold underlined font* shows known or predicted exon start sites; exon numbers refer to the human *CES1* gene (see Langmann et al. 1997). (Color figure online)

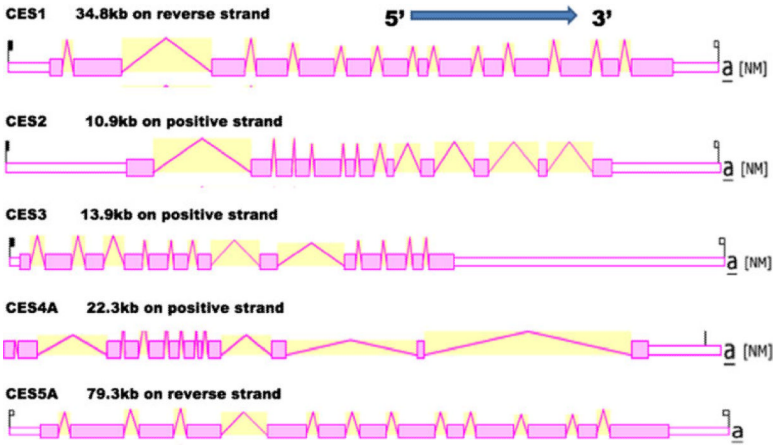

Fig. 2.

Gene structures and major isoforms for human *CES1*, *CES2*, *CES3*, *CES4A*, and *CES5A* genes. Derived from AceView website http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/ (Thierry-Mieg and Thierry-Mieg 2006). Mature isoform variants (a) are shown with capped 5′ and 3′ ends for the predicted mRNA sequences. Exons are in *solid color*. 5′ and 3′ untranslated regions of the genes are shown as *open boxes*. Introns are shown as a line. The 5′ → 3′ transcription directions are shown. a refers to the major transcript isoform for each human *CES* gene. Note that each *CES* gene structure is drawn to a different scale and that the respective gene sizes are shown: *CES1*, 34.8 kb; *CES2*, 10.9 kb; *CES3*, 13.9 kb; *CES4A*, 22.3 kb; and *CES5A*, 79.3 kb. (Color figure online)

Mamm Genome. Author manuscript; available in PMC 2011 October 8.

Table 1

Human *CES1*, *CES1P1*, *CES2*, *CES3*, *CES4A*, and *CES5A* genes and subunits

Human *CES* gene	Chromosome 16 coordinates	Gene size (bp)	Exons strand	Subunit MW	Amino acids	GenBank ID	Other gene names	Expression tissues (relative level of gene expression)	NCBI RefSeq transcript	UNIPROT ID
CES1	54,394,465–54,424,468	30,004	14 −ve	62,521	567	L07765	*hCE-1, CES1A1, HU1, EST1*	liver, lung, others [×3.8]	NM_001025195	P23141
CES1P1	55,794,511–55,808,824	14,314	6 +ve	ps	ps	AF106005	*CES4*	pseudogene	NR_003276	
CES2	65,527,040–65,535,426	8,387	12 +ve	61,807	559	BC032095	*CE-2, HU2, hCE-2*	brain, kidney, intestine [×4.5]	NM_003869	O00748
CES3	65,552,712–65,564,450	11,739	13 +ve	62,282	571	BC053670	*ES31, CE3*	colon, brain, others [×0.5]	NM_024922	Q9H6X7
CES4A	65,580,177–65,600,543	20,367	14 +ve	60,366	561	BC166638	*ESTHL, CES8, CE5*	brain, lung, kidney [×0.7]	NM_173815	Q5XG92
CES5A	54,437,867–54,466,634	28,768	13 −ve	63,936	575	BC039073	*CES7, CE4*	brain, lung, testis [×0.1]	NM_001143685	Q6NT32

Human *CES* gene	Human *CES* transcript isoform names	Gene size (bp)	Exons strand	Subunit MW	Amino acids	GenBank ID	Other names for human CES isoforms	AceView[a] human CES isoform name	NCBI RefSeq transcript	Transcript length (bp)
CES1	*CES1_AB119997*	30,380	14 −ve	62,592	568	AB119997	*CES1A1*	CES1, variant aApr07	NM_001025195	2,084
	CES1_AB119996	30,380	14 −ve	62,521	**567**	AB119996	*CES1A2*	CES1, variant bApr07	NM_001025194	2,081
	CES1_AK290623	30,310	14 −ve	62,393	566	AK290623	*CES1A3*	CES1, variant cApr07	NM_001266	2,007
CES2[d]	*CES2_BC032095*	10,890	12 +ve	68,899	**559**	BC032095	*CES2A1*	CES2, variant aApr07	NM_003869	4,177
	CES2_AL713761	10,660	12 +ve	67,051	607	AL713761	*CES2A2*	CES2, variant bApr07	NM_198061	3,901
	CES2_AK095522	10,590	12 +ve	61,566	560	AK095522	*CES2A3*	CES2, variant cApr07	NM_003869	4,140
CES3	*CES3_AY358609*	13,920	13 +ve	62,282	**571**	AY358609	*CES3A1*	COesterase.1, variant aApr07	NM_024922	3,894
	CES3_BC053670	12,160	13 +ve	61,967	568	BC053670	*CES3A2*	COesterase. 1, variant bApr07	BC053670[b]	2,123
CES4A	*CES4A_BC166638*	20,367	14 +ve	60,366	**561**	BC166638	*CES4A1*	*c*	NM_173815	2,135
CES5A	*CES5A_BC069501*	29,217	13 −ve	63,926	**575**	BC069501	*CES5A1*	CES7, variant aApr07	NM_001143685.1	2,285
	CES5A_BC069548	29,217	12 −ve	58,201	525	BC069548	*CES5A2*	CES7, variant bApr07	NM_145024	2,135

RefSeq, GenBank, and UNIPROT IDs provide the sources for the gene and protein sequences; the relative gene expression level for human CES genes in comparison with the expression of an average human gene is given in brackets. Gene sizes are given as base pairs of nucleotides. CES isoform sequences aligned in Fig. 1 are bold

ps pseudogene (*CES1P1*), *+ve* and *−ve* transcription strand direction

[a] http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/

[b] GenBank ID number

[c] No current AceView isoform name available

[d] The human CES2_BC032095 isoform transcript contains multiple transcription start sites with the shorter CES2 sequence (559 residues) previously reported (Pindel et al. 1997; Schewer et al. 1997)

Table 2

Mouse *Ces* genes and subumts

Mouse CES gene (proposed)	Chr 8 coordinates	Gene size (bp)	Exons Strand[a]	Subunit MW	Amino acids	GenBank ID	MGI ID_YZ	Current MGI symbol_YZ	Current gene symbols	NCBI transcript	Vega ID	Ensembl ID	UNIPROT ID	Tissue expression (relative)[b]
Ces1a	95,544,116–95,572,091	27,979	14 −ve	61,744	563	BC089371	MGI:3648919	Gm4976	*EG244595*	NM_001013764	None	ENSMUSG 00000071047	Q5FWH4	Fetal liver [0.08]
Ces1b	95,580,789–95,603,815	23,027	13 −ve	62,197	567	*NM_001081372	MGI:3779470	Gm5158	*CesN*	NM_001081372	None	ENSMUSG 00000078964		Liver [×2.0]
Ces1c	95,622,914–95,655,182	32,268	13 −ve	61,172	554	BC028907	MGI:95420	Es1	*Es1, Ces-N*	NM_007954	ENSMUSG 00000024453	ENSMUSG 00000057400	P23953	Liver [×2.0]
Ces1d	95,690,157–95,721,618	31,462	14 −ve	61,788	565	BC019198	MGI:2148202	Ces3	*Ces3*	NM_053200	ENSMUSG 00000024539	ENSMUSG 00000056973	Q8VCT4	Tongue, liver [−2.2]
Ces1e	95,725,306–95,753,320	28,015	14 −ve	61,582	562	BC019208	MGI:95432	Es22	*Es22*	NM_133660	ENSMUSG 00000024532	ENSMUSG 00000061959	Q64176	Liver, kidney [0.4]
Ces1f	95,780,331–95,803,599	23,269	14 −ve	61,698	561	BC013479	MGI:234564	AU018778	*CesML1, TGH-2*	NM_144930	ENSMUSG 00000024519	ENSMUSG 00000031725	Q91WU0	Tongue, kidney [2.6]
Ces1g	95,826,807–95,861,053	34,247	14 −ve	62,680	565	BC021150	MGI:88378	Ces1	*Ces1*	NM_021456	ENSMUSG 00000024535	ENSMUSG 00000057074	Q3UW56	Tongue, kidney [2.6]
Ces1h	95,875,926–95,903,624	27,699	14 −ve	62,087	562	AK009689	MGI:75704	2310039D24Rik	*AK009689*	XM_134476	ENSMUSG 00000033579	ENSMUSG 00000074156		Tongue, kidney [2.6]
Ces2a	107,257,972–107,265,313	7,342	12 +ve	61,940	558	BC024491	MGI:2142491	Ces6	*Ces6*	NM_133960	OTTMUSG 00000027410	ENSMUSG 00000055730	Q8QZR3	Liver, colon [×1.0]
Ces2b	107,355,572–107,362,353	6,782	12 +ve	61,927	556	BC015286	MGI:2448547	BC015286	*BC015286*	NM_198172	OTTMUSG 00000027467	ENSMUSG 00000050097	Q6PDB7	Kidney, colon [0.1]
Ces2c	107,371,033–107,378,161	7,129	12 +ve	62,470	561	BC031170	MGI:2389505	Ces2	*Ces2*	NM_145603	OTTMUSG 00000027466	ENSMUSG 00000061825	Q91WG0	Kidney, colon [1.2]
Ces2d-ps	107,391,388–107,397,764	3,762	6 +ve			BC034182	MGI:3704319	Gm9756		XR_002069	None	ENSMUSG 00000031884		Pseudogene
Ces2e	107,450,221–107,457,611	7,391	12 +ve	62,735	560	BC055062	MGI:2443170	Ces5	*Ces5*	NM_172759	None	ENSMUSG 00000031886	Q8BK48	Liver, intestine [0.6]
Ces2f	107,471,256–107,479,862	7,335	12 +ve	62,707	561	BC117742	MGI:1919153	2310038E17Rik		NM_001079865	None	ENSMUSG 00000062826	Q08ED5	Tongue, thymus [0.2]
Ces2g	107,485,688–107,492,328	6,771	10 +ve	52,731	478	BC027185	MGI:1919611	2210023G06Rik		NM_197999	None	ENSMUSG 00000031877		Kidney, stomach [0.7]
Ces2h	107,524,753–107,544,307	19,554					MGI:3648740	Gm5744		XM_488149	None	None		Not available
Ces3a	107,572,572–107,582,000	21,512	13 +ve	61,510	554	AK138932	MGI:102773	Es31	*Es31*	NM_198672	None	ENSMUSG 00000069922	Q63880	Liver, aorta [1.1]
Ces3b	107,607,670–107,617,468	9,799	14 +ve	63,007	568	BC019047	Gm4738	Es31L	*Es31L*	NM_144511	None	ENSMUSG 00000062181		Liver [0.5]
Ces4a	107,655,852–107,673,417	17,566	14 +ve	62,123	563	BC026374	BC026374	Ces8	*Ces8*	NM_146213	OTTMUSG 00000027469	ENSMUSG 00000060560		Skin [0.1]
Ces5a	96,038,095–96,059,607	21,512	13 +ve	64,167	575	AB186393	MGI:1915185	Ces7	*Ces7*	NM_001003951	None	ENSMUSG 00000058019	Q8R0W5	Prostate [0.03]

RefSeq, GenBank, UNIPROT, MGI, Vega, and Ensembl IDs provide the sources for the gene and protein sequences; gene sizes are given as base pairs of nucleotides

http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/

ps pseudogene (*Ces2d-ps*)

[a] +ve and −ve = transcription strand

[b] The relative gene expression level for mouse *Ces* genes in comparison with the expression of an average mouse gene is given in brackets

Table 3

Rat *Ces* genes and subunits

Rat CES gene (proposed)	Chromosomes 19 (and 1) coordinates	Gene size (bp)	Exons strand[a]	Subunit MW	Amino acids	GenBank ID	RGD ID	Ortholog	Current gene symbols	NCBI RefSeq ID	Ensembl transcript ID	UNIPROT ID	Tissue expression [relative]
Ces1a	19:15,025,350–15,051,534	26,185	14 +ve	62,362	563		RGD:1583671	Mouse Gm4976	*LOC679817*	XM_001054575	ENSRNOT 00000060929	D4AA05	[0.01]
Ces1c	19:14,981,539–15,021,040	39,502	14 +ve	60,501	550	BC088251	RGD:2571	Mouse Es1	*Es1*	NM_017004	ENSRNOT 00000024622	P10959	Liver [0.2]
Ces1d	19:14,928,590–14,966,890	38,301	14 +ve	62,150	565	BC061789	RGD:70896	Mouse Ces3	*Ces3*	NM_133295	ENSRNOT 00000021812	P16303	Liver, lung [0.4]
Ces1e	19:14,887,969–14,924,191	36,223	14 +ve	61,715	561	X81395	RGD:621508	Mouse Es22	*Ces1, Es22*	NM_031565	ENSRNOT 00000020775	Q924V9	Liver [0.1]
Ces1f	19:14,849,955–14,876,723	26,769	14 +ve	62,495	561	BC128711	RGD:1642419	None specified	*LOC100125372*	NM_001103359	ENSRNOT 00000024187	Q64573	Kidney, liver [0.1]
Ces2a	19:37,855–44,723	6,869	13 −ve	61,802	558	AY834877	RGD:708353	Mouse Ces6	*Ces6*	NM_144743	ENSRNOT 00000015451	Q8K3RO	Liver [0.05]
Ces2c	1:267,887,436–267,894,795	7,360	12+ve	62,170	561	AB010632	RGD:621510	Mouse Ces2	*Ces2l*	NM_133586	ENSRNOT 00000045656	O70631, O70177	Brain, liver [0.1]
Ces2e	19:65,698–80,142	14,445	12 +ve	62,410	557	D50580	RGD:621563	Mouse Ces5	*Ces5*	NM_001100477	ENSRNOT 00000015724	O35535	Liver [0.01]
Ces2g	19:34,883,500–34,890,289	6,790	12+ve	62,909	560	CH473972	RGD:1308358	Mouse 2210023G05Rik	*2210023G05Rik*		ENSRNOT 00000048385	D3ZXQ0	Kidney, liver [0.06]
Ces2h	19:34,910,987–34,925,261	14,275	12+ve	62,280	557	BC107806	RGD:1560889	Gm5744	*Ces2*	NM_001044258	ENSRNOT 00000019072	Q32Q55	Intestine [0.08]
Ces2i	1:267,807,848–267,815,235	7,388	11 +ve	62,072	559	XM212849	RGD:1565045	Mouse Ces2	*RGD1565045*	XM_001074128	ENSRNOT 00000015997	D3ZE31	Not available
Ces2j	19:215,376–222,512	7,137	12 +ve	61,795	556		RGD:1591368	Mouse Ces2	*LOC685645*	XM_001074128	ENSRNOT 00000061734	D3ZP14	[0.01]
Ces3a	19:34,929,247–34,937,264	8,018	14 +ve	62,393	563		RGD:1588734	Human CES3			ENSRNOT 00000040499		Not available
Ces4a	19:34,948,579–34,965,647	17,069	14 +ve	63,446	563		RGD:1307418	Mouse Ces8	*Ces8*	NM_001106176	ENSRNOT 00000019169	D4AE76	[0.01]
Ces5a	19:11,910,831–11,938,412	27,582	11 +ve	64,401	575	AF479659	RGD:1549717	Mouse Ces7	*Ces7*	NM_001012056	ENSRNOT 00000049452	Q5GRG2	[0.01]

RefSeq, GenBank, UNIPROT, RGD, Vega, and Ensembl IDs provide the sources for the gene and protein sequences; gene sizes are given as base pairs of nucleotides; the relative gene expression level for rat *Ces* genes in comparison with the expression of an average rat gene is given in brackets

http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/

[a]+ve and −ve = transcription strand direction

Table 4

Functions and substrates for human *CES* and mouse and rat *Ces* genes and enzymes

Mammal	*CES* (*Ces*) gene	Current gene symbol(s)	Substrates and function (hydrolysis or detoxification)
Human	*CES1*	*CES1, hCE-1, CES1A1, HU1*	Heroin, cocaine[1–3], methyl phenidate[4], temocapril[5], CPT-11[6], flurbiprofen[7]
		CES1	Fatty acid ethyl ester synthase[8], sarin[9], ciclesonide[10], cholesteryl ester hydrolase[11], triacylglycerol hydrolase[11]
	CES2	*CES2, hCE-2, HU2*	Procaine[5], heroin, cocaine[1–3], temacapril[5], CPT-11,6 flurbiprofen[7], doxazolidine[12]
	CES3	*CES3*	CPT-11[6]
Mouse	*Ces1c*	*Es1, Ces-N*	Lung surfactant convertase[13], CPT-11[14]
	Ces1d	*Ces3*	Triacylglycerol hydrolase[15]
	Ces1e	*Es22, egasyn*	β-glucuronidase binding in the liver endoplasmic reticulum[16], retinyl ester hydrolase[26]
	Ces1f	*CesML1, TGH-2*	Triacylglycerol hydrolase[27], monoacylglycerol hydrolase[27], cholesteryl ester hydrolase[27], phospholipase[27]
	Ces1g	*Ces1*	Lipid metabolism[17]
	Ces2c	*Ces2*	Inducible liver acylcarnitine hydrolase[18]
Rat	*Ces1c*	*Es1*	Retinyl palmitate[19]
	Ces1d	*Ces3*	Cholesterol ester hydrolase[20], triacylglycerol hydrolase[27], retinyl ester hydrolase[28]
	Ces1e	*ES-3*	β-glucuronidase binding in the liver endoplasmic reticulum[21]
	Ces2a	*Ces6*	Intestinal first pass metabolism[22]
	Ces2c	*Ces2*	Inducible liver acylcarnitine hydrolase[18], intestinal first pass metabolism[22]
	Ces2e	*Ces5*	Intestinal first pass metabolism[22]
Cat	*CES5A*	*CES7, cauxin*	3-Methylbutanol-cysteinylglycine hydrolysis in urine releasing pheromone[23]
Rat, sheep	*CES5A*	*CES7, cauxin*	Lipid transfer reactions in epididymis[24]

[1] Pindel et al. 1997,

[2] Bencharit et al. 2003,

[3] Satoh and Hosokawa 2006,

[4] Sun et al. 2004,

[5] Takai et al. 1997,

[6] Humerickhouse et al. 2000, Xu et al. 2002, Ohtsuka et al. 2003, Morton et al. 2005,

[7] flurbiprofen derivatives serve as substrates, Imai 2006, Taketani et al. 2007, Hosokawa 2008,

[8] Diczfalusy et al. 2001,

*9*Hemmert et al. 2010,

*10*Mutch et al. 2007,

*11*Becker et al. 1994,

*12*Barthel et al. 2008,

*13*Krishnasamy et al. 1998, Ruppert et al. 2006,

*14*Morton et al. 2005,

*15*Dolinsky et al. 2005,

*16*Ovnic et al. 1991,

*17*Ellingham et al. 1998, Ko et al. 2009,

*18*Furihata et al. 2003,

*19*Sanghani et al. 2002,

*20*Ghosh et al. 1995, Okazaki et al. 2008,

*21*Robbi and Beaufay 1994,

*22*Masaki et al. 2007,

*23*Miyazaki et al. 2006,

*24*Ecroyd et al. 2006, Zhang et al. 2009,

*25*Gilham et al. 2005,

*26*Schreiber et al. 2009,

*27*Lehner and Vance 1999,

*28*Okazaki et al. 2006,

*29*Linke et al. 2005