Comparative Biochemistry and Physiology, Part D 4 (2009) 54–65

Contents lists available at ScienceDirect

Comparative Biochemistry and Physiology, Part D

journal homepage: www.elsevier.com/locate/cbpd

Horse carboxylesterases: Evidence for six *CES1* and four families of *CES* genes on chromosome 3

Roger S. Holmes [a,b,c,*], Laura A. Cox [a,b], John L. VandeBerg [a,b]

[a] Department of Genetics, Southwest Foundation for Biomedical Research, San Antonio, TX, USA
[b] Southwest National Primate Research Center, Southwest Foundation for Biomedical Research, San Antonio, TX, USA
[c] School of Biomolecular and Physical Sciences, Griffith University, Nathan, QLD, Australia

ARTICLE INFO

Article history:
Received 24 September 2008
Received in revised form 31 October 2008
Accepted 31 October 2008
Available online 5 November 2008

Keywords:
Horse
Amino acid sequence
Carboxylesterase
Evolution
Gene duplication

ABSTRACT

Carboxylesterases (CES) are responsible for the detoxification of a wide range of drugs and xenobiotics, and may contribute to cholesterol, fatty acid and lung surfactant metabolism. In this study, *in silico* methods were used to predict the amino acid sequences, secondary and tertiary structures, and gene locations for horse *CES* genes and encoded proteins, using data from the recently completed horse genome project. Evidence was obtained for six *CES1* genes closely localised on horse chromosome 3, for which the predicted *CES1* gene products are ≥74% identical. The horse genome also showed evidence for three other *CES* gene classes: *CES5*, located in tandem with the *CES1* gene cluster; and *CES2* and *CES3*, located more than 9 million base pairs downstream on chromosome 3. Horse *CES2*, *CES3* and *CES5* gene products shared 42–46% identity with each other, and with the *CES1* protein subunits. Sequence alignments of these enzymes demonstrated key enzyme and family specific CES protein sequences reported for human CES1, CES2, CES3 and CES5. In addition, predicted secondary and tertiary structures for horse CES1, CES2, CES3 and CES5 subunits showed extensive conservation with human CES1. Phylogenetic analyses demonstrated the relationships and potential evolutionary origins of the horse CES sequences with previously reported sequences for human and other mammalian *CES* gene products. Several *CES1* gene duplication events have apparently occurred following the appearance of the 'dawn' horse ~55 million years ago.

1. Introduction

Carboxylesterases (CES; E.C.3.1.1.1) catalyse hydrolytic and trans-esterification reactions using a broad range of substrates, including xenobiotics, anticancer pro-drugs, narcotics, clinical drugs and pro-herbicidal esters (Redinbo and Potter 2005; Gershater et al., 2006; Satoh and Hosokawa, 2006). CES also detoxifies organophosphates, carbamate compounds and insecticides (Leinweber 1987), catalyses several cholesterol and fatty acid metabolic reactions (Hosokawa et al., 2007) and the conversion of alveolar surfactant in lung (Ruppert et al., 2006); and has been linked with the assembly of low density lipoprotein particles in liver (Wang et al., 2007).

Five families of mammalian CES have been reported (Holmes et al., 2008a) including CES1, the major liver enzyme (Shibita et al., 1993); CES2, the major intestinal enzyme (Schewer et al., 1997); CES3, expressed in liver, colon and brain (Sanghani et al., 2004); CES5, a major urinary protein of the domestic cat (Miyazaki et al., 2003; Holmes et al., 2008b); and CES6, a predicted CES-like enzyme in brain (Clark et al., 2003). Three-dimensional structural analyses of human CES1 have clarified the structure–function relationships for this enzyme, and the identification of three ligand binding sites, including the promiscuous active site, 'side door' and 'Z-site', where substrates, fatty acids and cholesterol analogues respectively, are bound; and a 'product releasing' residue (Bencharit et al., 2003, 2006; Fleming et al., 2005).

Structures for several human and animal *CES* genes have been determined, including human (see Ghosh, 2000; Marsh et al., 2004) and rodent *CES1* and *CES2* 'like' genes (see Hosokawa et al., 2007). Moreover, following the release of a number of mammalian genome sequences, predicted *CES* gene structures have been described for five classes of *CES* genes in several mammals and other animal species (Holmes et al., 2008a,b, in press-a,b). Recently, the horse (*Equus caballus*) genome sequence has been reported (Horse Genome Project, 2008) enabling *in silico* interrogation and analyses of horse genes and proteins to be undertaken. This paper reports the predicted gene and amino acid sequences; predicted secondary and tertiary structures for multiple horse CES1 protein subunits and CES2, CES3 and CES5 protein subunits; and describes the structural, phylogenetic and evolutionary relationships for these enzymes.

* Corresponding author. Department of Genetics, Southwest National Primate Research Center, Southwest Foundation for Biomedical Research, San Antonio, TX, 78227, USA. Tel.: +1 210 258 9687; fax: +1 210 258 9600.
E-mail address: rholmes@sfbrgenetics.org (R.S. Holmes).

1744-117X/$ – see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.cbd.2008.10.004

Even though horse liver CES was one of the first enzymes subjected to large scale purification (Burch, 1954) and has been biochemically characterized (Stoops et al., 1975; Inkerman et al., 1975), there are no previous reports of protein and genomic structures and sequences for horse CES for any of the mammalian CES classes. CES has been extensively investigated in other mammals and shown to *serve a range of* metabolic (Satoh and Hosokawa 2006; Redinbo and Potter 2005) and biomedical roles (Pindel et al., 1997; Xu et al., 2002; Imai, 2006; Mutch et al., 2007; Wang et al., 2007).

2. Methods

2.1. In silico horse CES gene and protein identification

BLAST (*B*asic *L*ocal *A*lignment *S*earch *T*ool) studies were undertaken using web tools from the National Center for Biotechnology Information (NCBI) (http://blast.ncbi.nlm.nih.gov/Blast.cgi) (Altschul et al., 1997). Protein BLAST analyses used the human CES1 amino acid sequence (see Table 1; Fig. 1) to examine the non-redundant protein sequences database available for the horse genome (Horse Genome Project, 2008) using the blastp algorithm. This procedure produced 25 BLAST 'hits' which were individually examined and retained in FASTA format, and a record kept of the sequences for predicted mRNAs and encoded forms of horse CES and related proteins. These records were derived from annotated genomic sequences using the gene prediction method: GNOMON and predicted sequences with high similarity scores for human CES1 were further examined. Nine CES-like reference sequences were obtained, including six predicted as being CES1-like, and three predicted as CES2-like, CES3-like and CES5-like (see Table 1).

BLAT (*B*LAST-*L*ike *A*lignment *T*ool) *in silico* analyses were subsequently undertaken for each of the predicted horse CES amino acid sequences using the UC Santa Cruz web browser [http://genome.ucsc.edu/cgi-bin/hgBlat] (Kent et al., 2003) with the default settings to obtain the predicted locations for each of the horse *CES* genes, including predicted exon boundary locations and gene sizes. Sequences for other known human *CES* gene products, including *CES2*, *CES3*, *CES5* and *CES6* (Table 1), were also used in BLAST analyses to examine the horse non-redundant protein sequence database. With the exception of *CES6*, predicted mRNA and encoded protein sequences were obtained for each of the corresponding horse *CES* genes, and BLAT analyses revealed predicted gene locations and exon boundaries for each of the horse *CES* genes. BLAT analyses were also undertaken of the horse genome using the UC Santa Cruz web browser to obtain predicted nucleotide sequences for exons 13 and 14 and intron 13 for each of the six *CES1*-like genes (designated *CES1.1; CES1.2; CES1.3; CES1.4; CES1.5; and CES1.6*) using the derived amino acid sequences to interrogate the horse genome.

2.2. Predicted structures and properties for horse CES gene products

Predicted secondary and tertiary structures for horse CES1 subunits (1.1–1.6), CES2, CES3 and CES5 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 (Kopp and Schwede 2004). The reported tertiary structure (2.0 Å resolution) for the human CES1 Coenzyme A complex (Bencharit et al., 2003, 2006; Fleming et al., 2005) served as the reference for obtaining the predicted horse CES tertiary structures, with a modeling range of residues 21–551 for the horse CES1 subunits; residues 29–539 for horse CES2; residues 32–550 for horse CES3; and residues 31–540 for horse CES5. Theoretical isoelectric points and molecular weights for horse CES 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 horse CES sequences (Emanuelsson et al., 2007).

2.3. 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; Larkin et al., 2007) [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,

Table 1
Horse, other mammalian, Xenopus and zebrafish *CES* genes

CES Gene	[a]GenBank mRNA [b]NCBI Locus [c]UNIPROT ID	No. of amino acids	Chromosome location	Strand	Gene size kbs	Predicted subunit *p*I	Predicted subunit MW
Horse CES1.1	[b]XP1491160	565	3:7,903,234–7,982,438	Negative	22	5.5	61,486
Horse CES1.2	[b]XP1491576	565	3:8,000,851–8,032,477	Negative	79.2	5.6	61,556
Horse CES1.3	[b]XP1491752	565	3:8,048,349–8,079,872	Negative	31.6	5.5	61,935
Horse CES1.4	[b]XP1491878	565	3:8,106,399–8,137,176	Negative	31.5	5.6	61,742
Horse CES1.5	[b]XP1491878	565	3:8,167,845–8,199,077	Negative	30.8	6.5	61,037
Horse CES1.6	[b]XP1915508	566	3:8,233,589–8,266,613	Negative	31.9	5.5	64,050
Horse CES2	[b]XP19115822	559	3:17,402,324–17,411,849	Positive	8.9	5.5	61,670
Horse CES3	[b]XP1496251	571	3:17,417,834–17,428,268	Positive	10.4	6.2	62,719
Horse CES5	[b]XP1493477	575	3: 8,287,811–8,319,572	Negative	31.8	5.6	63,859
Human CES1	[a]L07765 [c]P23141	567	16:54,394,266–54,424,4894	Negative	30	6.2	62,521
Baboon CES1	[a]FJ1471785	567				6.1	62,483
Mouse CES1	[a]Y12887 [c]Q8VCC2	565	8:95,826,807–95,861,053	Negative	34.2	5.6	62,680
Bovine CES1.1	[a]BC102781	565	18: 24,344,904–24,371,523	Negative	26.6	6.3	61,723
Human CES2	[a]BX538086 [c]O00748	559	16:65,527,040–65,535,426	Positive	8.4	5.7	61,807
Human CES3	[a]AK025389 [c]Q6UWW8	571	16:65,552,712–65,564,450	Positive	11.7	5.4	62,282
Human CES5	[a]AK056109 [c]Q6NT32	575	16:54,437,867–54,466,634	Negative	28.8	6	63,926
Human CES6	[a]FLJ37464 [c]Q5XG92	561	16:65,580,177–65,600,543	Positive	20.4	9.4	63,529
Zebrafish CES	[a]BC091470 [c]Q1LUZ9	548	18: 16,714,030-16,721,981	Negative	7.9	6	60,297
Xenopus CES	[a]BC082503 [c]A1L2G7	557	scaffold170:71,809-89,385	Negative	17.6	5	61,707

*p*I refers to theoretically determined isoelectric point. kbs refers to kilobases of DNA nucleotides.
[a] Locus derived from a BLAST of the NCBI data base (http://blast.ncbi.nlm.nih.gov/Blast.cgi).
[b] GenBank ID (http://www.ncbi.nlm.nih.gov/Genbank/).
[c] UNIPROT ID of the CES protein using the SWISS-PROT Web Browser (http://au.expasy.org/).