Baboon Carboxylesterases 1 and 2: Sequences, Structures and Phylogenetic Relationships with Human and other Primate Carboxylesterases

Roger S. Holmes[1,2,3], Jeremy P. Glenn[1], John L. VandeBerg[1,2], Laura A. Cox[1,2,4]

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

4. Corresponding Author:

Laura A. Cox, Ph.D.
Department of Genetics
Southwest National Primate Research Center
Southwest Foundation for Biomedical Research
San Antonio, TX, USA 78227
Email: lcox@sfbrgenetics.org
Phone: 210-258-9687
Fax: 210-258-9600

Keywords: cDNA sequence; amino acid sequence; 3-D structure

Running Head: Carboxylesterases: sequences and phylogeny

Abstract

Background Carboxylesterase (CES) is predominantly responsible for the detoxification of a wide range of drugs and narcotics, and catalyze several reactions in cholesterol and fatty acid metabolism. Studies of the genetic and biochemical properties of primate CES may contribute to an improved understanding of human disease, including atherosclerosis, obesity and drug addiction, for which non-human primates serve as useful animal models.
Methods We cloned and sequenced baboon CES1 and CES2 and used in vitro and in silico methods to predict protein secondary and tertiary structures, and examined evolutionary relationships for these enzymes with other primate and mouse CES orthologs.
Results and Conclusions We found that baboon CES1 and CES2 proteins retained extensive similarity with human CES1 and CES2, shared key structural features reported for human CES1, and showed family specific sequences consistent with their multimeric and monomeric subunit structures respectively.

Introduction

Carboxylesterases (CES; E.C.3.1.1.1) exist as a family of enzymes which have been implicated in the catalysis of a broad range of hydrolytic and transesterification reactions [56]. Diverse substrates include xenobiotics (carboxyl esters, thioesters and aromatic amides), anticancer drugs such as CPT-11 and capecitabine, narcotics such as heroine and cocaine and clinical drugs such as lovastatin and lidocaine [50, 58]. CES detoxifies numerous organophosphate and carbamate compounds used as chemical weapons (sarin, tabun and soman) or insecticides (malathion), usually by 'suicide-inactivation' at the CES active site [2; 33]. The enzyme has also been associated with cholesterol and fatty acid metabolism through demonstrated activities for several reactions: cholesterol:ester hydrolase [21]; fatty acyl CoA hydrolase [62]; acyl CoA:cholesterol acyl transferase [3]; acyl carnitine hydrolase [26]; and fatty acyl: ethyl ester synthase [15], and has been linked with the assembly of low density lipoprotein particles in liver [65]. CES is predominantly localized in the endoplasmic reticulum (ER), has an N-terminal hydrophobic signal peptide consistent with a trafficking role through the ER and forms complexes with β-glucuronidase, consistent with a partnering role in phase II drug metabolism [47, 69].

Two major forms of human CES have been reported: CES1 is the major liver enzyme and also is found in lung epithelia, macrophages and other tissues [61]; and CES2, the major intestinal enzyme which is widely distributed in tissues

including liver, kidney, heart and skeletal muscle [32]. Three-dimensional structural analyses of several human CES1 complexes have shown that the active site is highly promiscuous enabling catalysis of many diverse reactions and using a wide range of substrates [5, 6, 36, 50, 58]. In addition to the active site, human CES1 has at least two ligand binding sites, the 'side-door' and 'Z-site', where fatty acids and cholesterol analogues respectively, are bound. The trimer-hexamer subunit structure for CES1 is proposed to play an important role in the regulation of catalysis through ligand binding, which may influence access to the active site by shifting this equilibrium [50]. Genomic structures for the genes encoding these enzymes have been determined: CES1 on chromosome 16 contains 14 exons and spans about 30 kb [9, 32, 61]; and CES2 also on chromosome 16 contains 15 exons and spans about 11 kb [10, 36, 59]. Their close proximity and homology (47% identity for human CES1 and CES2 cDNAs) imply that they arose from an ancestral gene duplication event [46, 56-58]. Three other CES genes (CES3, CES4 and CES7) have been also mapped on chromosome 16 of the human genome and products identified [39, 55, 67].

This paper reports the cDNA and deduced amino acid sequences, subunit structures and predicted secondary and tertiary structures for baboon CES1 and CES2, and describes the structural, phylogenetic and evolutionary relationships for these enzymes from several primate species, including humans. The baboon and other non-human primate species are used as animal models for studying several human diseases [54], including atherosclerosis and related arterial diseases [7, 35], obesity [13] and drug addiction and toxicity [40, 52]. Given the major role of CES in the metabolism of drugs and narcotics [50, 58], and in catalyzing several reactions of cholesterol and fatty acid metabolism [3, 14, 21, 26, 62], these and related studies may lead to a better understanding of genetic and biochemical factors contributing to these common diseases.

Methods

cDNA cloning and sequencing.

Total RNA was extracted from baboon liver. First strand cDNA synthesis from total RNA was performed with an Oligo (dT) Primer (Ambion, Austin,TX) and Superscript III Reverse Transcriptase (Invitrogen, Carlsbad,CA) according to the manufacturer's instructions. CES1 baboon cDNA was PCR amplified from one first strand cDNA sample using forward primer (5' – AAAACTGTCGCCCTTCCACG -3') and reverse (5' – TTCCCCAGCCACGGTAAGATGCCT -3') designed from rhesus sequence data [28, 51]. CES2 baboon cDNAs were PCR amplified as 3 overlapping fragments from each first strand sample using: fragment 1 (698 bp – 2532 bp) forward primer (5' – TTTGCTCAAGCGGTTCCTTC -3') and reverse primer (5' – TCATGTGCGGTGGCCTGATGTTCT -3'); fragment 2 (1171 bp – 2532 bp) forward primer (5' – AACCGAGACCAGCGAGCCGACCAT -3') and reverse primer (5' – TCATGTGCGGTGGCCTGATGTTCT -3'); and fragment 3 (2429 bp – 3862 bp) forward primer (5' – GCGGACTCCATGTTTGTGATCCCT -3') and reverse primer (5' – ACTTAGGTGTGGGCAACATTCTTC -3') designed from human sequence data [28]. Baboon CES1 and CES2 cDNAs were sequenced from the PCR fragments and a tiling path of sequencing primers for each gene was constructed (Table 1). cDNA samples for each baboon were sequenced using these primers and Big Dye v3.1 chemistry (Applied Biosystems Cat. No. 4337455). Sequencing products were purified using Exonuclease I (10ul/ul, USB Cat. No. 70073Z) and Shrimp Alkaline Phosphatase (SAP) (1u/ul, USB Cat. No. 70092Z) and analyzed on an automated sequencer (Applied Biosystems 3100) using Sequence Analysis software (Applied Biosystems 3100). Sequence data were imported into Sequencher (Gene Codes, Inc.) for alignment and identification of polymorphisms.

Phylogenetic Studies and Sequence Divergence

Phylogenetic trees were constructed using an amino acid alignment from a ClustalW-derived alignment of primate and mouse CES protein sequences, obtained with default settings and corrected for multiple substitutions [12, 29, 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 terminus, were excluded prior to phylogenetic analysis yielding an alignment of 528 residues. 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. Table 2 provides a summary of the Genbank (http://www.ncbi.nlm.nih.gov/Genbank/), UniProtKB/Swiss-Prot (http://au.expasy.org/sprot/) and predicted in silico CES sequences [28] used in alignment studies. Also included are the baboon CES1 and CES2 sequences reported here. The extent of divergence for mammalian CES1 and CES2 amino acid sequences was determined using the SIM-Alignment tool for Protein Sequences [http://au.expasy.org/tools/sim-prot.html].

In silico CES gene and gene product identification.

BLAT (BLAST-Like Alignment Tool) in silico studies were undertaken using the UC Santa Cruz [http://genome.ucsc.edu/cgi-bin/hgBlat] web site with the default settings [28]. UniProtKB/Swiss-Prot Database [http://au.expasy.org] and GenBank [http://www.ncbi.nlm.nih.gov/Genbank/] sequences (see Table 2) were used to interrogate the human, chimp, rhesus, mouse, rat, cow, dog and cat genome sequences [28] and gene locations were observed for each CES sequence examined for those regions showing identity with the respective CES gene products.

Purification of CES1 and CES2 and native MW studies

Baboon liver samples stored at -80°C were extracted into 50mM Tris-HCl buffer pH 8.0 containing 0.1% Triton-X 100 (buffer A) (20% w/v) using an Ultra-Turrax homogenizer and centrifuged at 4°C for 30 min at 15,000g. Extracts (~ 5μl) were then subjected to electrophoresis on Titan III cellulose acetate plates at 4°C using 50mM Tris-citrate buffer pH 7.0 at 150 V for 60 min. Edges of the plate were cut and histochemically stained for CES activity using α-napthyl acetate as substrate, Fast Blue RR salt and 1% agarose-overlay in 50mM Tris-HCl pH 8.0 buffer [27]. Cellulose acetate particles were scraped from the unstained plate for the CES1 and CES2 activity zones, and gently extracted into 0.25 mls of Buffer A and centrifuged at 2,000g for 10 min at 4°C. The supernatants containing isozymically purified CES1and CES2 were then subjected to gradient-PAGE (4-20%) in a Tris-Glycine pH 8.5 buffer at 150V for 4.45 hours at 4°C and histochemically stained for CES activity. Native MW standards were run simultaneously with the purified CES1 and CES2 samples and these lanes stained for protein with Coomassie Blue G-250 (0.25%w/v in a solution of 30% methanol/10% acetic acid and distilled water) and destained in the methanol/acetic acid solution. Native MWs for baboon liver CES1 and CES2 were determined from a linear plot of log10MW for the protein standards and migration distance.

Predicted Secondary and Tertiary Structures for Baboon CES Gene Products.

Predicted secondary structures for baboon CES1 and CES2 were obtained using the PSIPRED v2.5 web site tools provided by Brunel University [38, http://bioinf.cs.ucl.ac.uk/psipred/psiform.html]. Predicted tertiary structures for baboon CES1 and CES2 were obtained using the SWISS MODEL web tools [http://swissmodel.expasy.org/workspace/index.php?func=modelling_simple1] [24, 30, 45, 68]. The reported human CES1 tertiary structures [4, 5, 50] served as the reference for obtaining the predicted baboon CES1 and CES2 tertiary structures.

Results

Baboon liver CES1 and CES2 cDNA Sequences.

We isolated and sequenced cDNAs for baboon CES1 and CES2 and lodged the respective sequences with GenBank (GenBankxxxx and GenBankyyyy, respectively). Based on an alignment with the human CES1 sequence [9, 21, 32] (data not shown), the baboon CES1 cDNA sequence includes 21 bp of the 108 bp 5′ untranslated region (UTR), the entire coding sequence, and 47 bp of the 212 bp 3′UTR. We also sequenced 1980 bp of the baboon CES2 cDNA. An alignment of the baboon CES2 cDNA sequence with the reported human CES2 cDNA sequence [10, 55, 59, 61] (data not shown) showed that the sequence includes 57 bp of 1192 bp 5' UTR, the entire coding sequence and 237 bp of the 1085 bp 3' UTR. Alignments of baboon CES1 and CES2 cDNAs with the corresponding sequences for human CES1 and CES2 showed 95% and 92% sequence identities respectively with both the non-coding and coding regions of the genes.

Alignments of baboon CES1 and CES2 with human and other primate CES1 and CES2 amino acid sequences.

The deduced amino acid sequences for baboon liver CES1 and CES2 are shown in Figure 1 together with previously reported sequences for human CES1 [9, 21, 32] and human CES2 [10, 55, 59, 61], together with deduced sequences for orangutan CES1 [64], chimp CES1 and CES2, and rhesus CES1 and CES2. The latter four sequences were obtained by in silico interrogation of the chimp and rhesus genomes using human CES1, human CES2, baboon CES1 and baboon CES2 sequences, respectively [28]. Alignments of baboon CES1 with human CES1 and of baboon CES2 with human CES2 showed 94% and 90% sequence identities respectively, while other primate CES1 amino acid sequences are 93% or more identical with human CES1 whereas other primate CES2 sequences are 86% or more identical with human CES2 (Table 3). The deduced amino acid sequence for baboon CES2 was two residues longer (Met310 and Lys546 within the 561 amino acid sequence) than for human CES2 (559 residues). Human CES1 occurs as two major isoforms containing 567 and 568 residues for isoforms 1 and 2 respectively [3, 9], in comparison with the reported baboon CES1 sequence which contains 567 amino acid residues. Since the human CES-isoform 1 is the sequence used to describe the tertiary structure for this enzyme [4, 5, 17, 50], these residue numbers are cited in the following text for comparing amino acid sequences of several primate CES1 proteins. All of the primate CES forms examined were between 559-567 residues in length, of which 228 residues (~40%) were identical. Of particular interest are the key residues involved in

determining the catalytic, microlocalization and regulatory functions for these enzymes which will be dealt with in more detail below and in the discussion section.

Predicted Secondary and Tertiary Structures for Baboon and other Primate CES Gene Products.

Analyses of predicted secondary structures for human CES2, and for baboon and other primate CES1 and CES2 proteins were compared with the previously reported secondary structure for human CES1 [4, 5, 50] (Figure 1). The αhelix-βsheet predominant CES structure is readily apparent with a high degree of similarity in secondary structures for the CES gene products examined. Consistent structures were apparent near key residues or functional domains including the α-helix within the N-terminal signal peptide (residues 2-13 and 3-24 for human CES1 and CES2, respectively); the β-sheet and α-helix structures bordering the active site Ser228 (human CES1) and the active site and nearby Z site(Glu354/Gly356 respectively); the α-helix structures bordering the 'side door' site; and the α-helix located near the 'gate' at the CES1 and CES2 C-terminus end [4, 5, 50]. In addition, two random coil regions (residues 51-115 and 169-188 for human CES1) are retained for all forms of primate CES1 and CES2 which have been shown for human CES1 to contain 2 charge clamps sites: Lys79 +-Glu183; and Glu73-+Arg186); an N-glycosylation site at 79NAT; a second potential N-glycosylation site for all 4 forms of primate CES2 (111NMT for human CES2), and one of the disulfide bridges (87Cys –S-S-117Cys) reported for human CES1 [3, 4, 47] . Predicted 3-D structures for baboon CES1 and CES2 (Figure 2) showed a high degree of similarity with human CES1 [4, 5, 50]. The rainbow based color code (red for carboxyl-end and deep blue for amino terminus end) illustrated the high degree of conservation observed for both baboon CES1 and CES2 secondary and tertiary structures despite having <45% sequence identity.

Subunit Structures for Baboon Liver CES1 and CES2.

Baboon liver CES1 exhibited gradient PAGE migration properties consistent with being a mixture of two multi-subunit structures, trimers (~170kD) and hexamers (340kD), whereas liver and intestine CES2 behaved as a monomer, based upon subunit MWs of ~62kD for both enzymes (Figure 3) [9, 10]. This is consistent with the results reported for human CES1 and CES2 as trimers/hexamers and monomers, respectively [46]. Baboon intestine CES1 exhibited similar properties to liver CES1 but with two higher MW zones indicating further aggregation of subunits for this enzyme. A monomeric subunit structure for baboon CES2 is further supported by the amino acid replacements observed at the subunit-subunit binding sites of human CES1 [17] for the corresponding residues in human and baboon CES2 (Figure 1). Two charge clamp residues identified for human CES1 [17] have been replaced by amino acids in both human and baboon CES2 which would interfere with monomer-monomer binding: Glu183 in human and baboon CES1 binding to Lys78 in the adjacent subunit is substituted by Lys190 in human and baboon CES2, and in other primate CES2 proteins examined (Figure 1); and Arg193 binding to Glu72 in the adjacent subunit has been substituted by Thr200 (Fig.1) for the four primate CES2 proteins. In addition, 79Asn in human and baboon CES1 has been replaced by 87Asp (Fig.1) in CES2 for the 4 primates, preventing carbohydrate binding to the amide group and sialic acid stacking which is essential for trimer-hexamer formation [17]. The loss of two charge clamps and the loss of a carbohydrate binding site within CES2 may contribute significantly to the disruption of monomer-monomer binding leading to the observed single subunit structure for human and baboon CES2.

Phylogeny and Divergence of Mammalian CES1 and CES2 Sequences.

A phylogenetic tree (Figure 4) was calculated by the progressive alignment of nine primate and two mouse CES1 and CES2 amino acid sequences which shows clustering into two main groups (or families) for the CES1 and CES2 'like' genes. The average amino acid sequence divergence rates for primate CES1 and CES2 were calculated using the known date for the common Catarrhine ancestor for these species (~25 million years ago [MY ago]) [23]: the average amino acid substitution rate in each case was 0.13 (± SEM) % and 0.24 (± SEM) % amino acid substitution per MYA (Table4). The average amino acid substitution rates for 20 CES1 and CES2 gene products among non-primate eutherian mammals were also calculated using the known date for the eutherian common ancestor (~ 100 MY ago) [66] as 0.14 (± SEM)% and 0.16 (± SEM) %, respectively (Table 4). Primate CES 2 genes are apparently undergoing a more rapid rate of amino acid sequence divergence than primate CES1 genes and non-primate CES1 and CES2 genes, which is further substantiated by comparing the percentage of sequence identities for human and chimp CES1 and CES2 with those for baboon and rhesus CES1 and CES2 (Table 3). In each case, a significantly lower percentage of identity was observed for baboon and rhesus CES2 with human and rhesus CES2 than for the same inter-species comparisons for primate CES1.

Gene Locations for Mammalian CES1 and CES2 Genes.

Table 2 summarizes the known or predicted CES1 and CES2 gene locations for 3 primate species and for 5 non-primate eutherian mammals based upon published reports for human [9, 10, 32, 59, 61], mouse [16, 18, 41] and rat CES1 and CES2 [19, 49], and BLAT interrogation of the genomes for human, chimp, rhesus, mouse, rat, cow, dog and cat, using

the UC Santa Cruz Gene Browser [28]. With the exception of rat CES2 (CES2 'like') and CES3 (CES1 'like') genes, which are located on chromosomes 1 and 19 respectively, CES1 and CES2 genes from the other 5 mammalian genomes examined were located on the same chromosome in each case. In addition, ten of eleven mammalian CES1 genes were transcribed on the negative strand, and nine of ten CES2 genes were transcribed on the positive strand.

Discussion

Three dimensional structures for human CES1 have been determined at high resolution (2.8Å) [4, 5, 50] and the enzyme shown to be divided into three functional domains: the catalytic domain contains the active site 'triad' and the carbohydrate binding site; the αβ domain provides the majority of the hydrophobic internal structure and assists in forming the trimeric subunit structure for this enzyme; and the regulatory domain which may regulate substrate binding, product release and the trimer-hexamer equilibrium. Three ligand binding sites have been described, designated as the active site, the side door and the Z-site. The side door apparently assists with the release of product (eg. fatty acids) following catalysis, whereas the Z-site is proposed to play a role in regulating catalysis following ligand binding by shifting the trimer-hexamer equilibrium towards trimer, and opening up the active site to substrate and subsequent catalysis [17]. Several key amino acid residues or sequences have been strictly conserved among primate CES1 and CES2 (Figure 1), which correlate with their proposed functions, based on the 3-D studies conducted on human CES1 [see 4, 5, 50] : the active site 'triad' (Ser221, Glu354 and His468) [12]; Gly356 or the Z-site, which binds cholesterol-like compounds; Cys95/Cys123 and Cys280/Cys291, the sites for disulfide bond formation [34]; and two microsomal targeting sequences, including the hydrophobic N-terminus signal peptides for CES1 (residues 1-18) and CES2 (residues 1-26) [63] and the C-terminal endoplasmic reticulum (ER) retention sequences His-X-Glu-Leu (HXEL), which functions in protein retrieval from the Golgi apparatus and in CES retention in the ER lumen [53].

There are other conserved amino acid residues or sequences which appear to be CES1 or CES2 specific among the primate sequences examined, and correlate with the functions proposed by the human CES1 tertiary structure studies [4, 5, 50]. The high-mannose N-glycosylation site reported for human CES1 (Asn79-Ala80-Thr81) [17] was retained for all primate CES1 sequences whereas another potential carbohydrate binding site was observed for all primate CES2 sequences at Asn111-X112-Thr113. A second potential carbohydrate binding site was also found for human and chimp CES2 at 276NLS (Figure 1). Given the roles described for N-glycosylation in maintaining human CES1 activity [31] and in stabilizing the N-terminal structure [17], it is likely that these potential sites for primate CES2 are also subject to N-glycosylation in vivo to assist with the stability for this enzyme. The charge clamps that perform key roles in maintaining the trimeric-hexameric subunit structures for human CES1 [17] are retained by baboon and rhesus CES1. Chimp and orangutan CES1, however, have retained only one of these clamps (Glu183 binding to Lys78 in the adjacent CES1 subunit), whereas the second charge clamp (Arg193 binding to Glu72 in the adjacent CES1 subunit) will not function as for human CES1 because of the Arg193 substitution by 193Pro. The monomeric subunit structure for baboon CES2 (Figure 3) is consistent with a previous report for human CES2 [46], and may be explained by the absence of key residues supporting two charge clamps previously reported for human CES1 [17]. The four primate CES2 sequences examined have undergone amino acid substitutions for those residues contributing to the human CES1 charge clamps: human CES1 Glu183 and Arg186 have been replaced for primate CES2 by amino acids that would not support charge clamp formation: Glu183 → Lys183; and Arg186 → Thr183 (Figure 1). It would appear that the respective multimeric and monomeric subunit structures for CES1 and CES2 have been retained for all primate species examined. This is likely to have a major influence on the kinetics and biochemical roles for these enzymes. Bencharit and coworkers [4, 5] have proposed that ligand binding to the CES1 'Z-site' shifts the trimer-hexamer equilibrium towards the trimer facilitating substrate binding and enzyme catalysis. They also proposed that ligand binding to this site may play a role in facilitating the hydrolysis of cholesterol esters and in the allosteric activation of esterase catalysis. As a monomer, primate CES2 would behave quite differently and may serve a distinct set of roles in drug and lipid metabolism in the body.

Human CES1 Met425 has been described as the most important residue in regulating the release of fatty acids following hydrolysis of cholesterol esters and serving as a 'gate' at the 'side door' of human CES1, where Phe426 serves as a switch' and Phe551 acts as an aromatic releasing residue also in the 'side door' region [4, 5, 50]. The 424VMF 'side door' amino acid residues (Figure 1) have been retained for all five primate CES1 sequences examined, whereas only the human CES1 425MF amino acids have been retained for the four primate CES2 sequences. This may reflect a change in the kinetics or specificity in regulating fatty acid or other acyl-product release for primate CES2 as compared with CES1. In addition, the key residue at the human CES1 'gate', 551Phe, has been conserved for all primate CES1 sequences examined, whereas another hydrophobic amino acid (human CES2 542Leu) has replaced the human CES1 551Phe residue in all four primate CES2 sequences. The significance of this change in 'gate' residue for primate CES2 remains to be determined however it may influence product release following acyl hydrolysis or transesterifcation. The deduced

amino acid sequence for baboon and rhesus CES2 was two residues longer (Met310 and Lys546 within the 561 amino acid sequence for baboon CES2) than for human and chimp CES2 (559 residues) (Figure 1). Human, chimp and baboon CES1 contained 567 amino acid residues whereas orangutan and rhesus CES1 were one residue shorter (566). The N-terminal microsomal signal peptides for human and baboon CES1, which facilitates retention of the enzyme within the ER [44, 53, 63], were identical in sequence between human CES1 isoform 2 and baboon CES1, and both contained an extra residue (Ala18) prior to Gly19 which is immediately prior to the cleavage site for human CES1 isoform 1. This is in contrast to most mammalian CES1 and CES2 products which have N-terminal signal peptides containing a bulky aromatic residue followed by a small neutral residue prior to the cleavage site [63].

There were three major predicted secondary structure differences observed between primate CES1 and CES2, involving longer helices for the four primate CES2 proteins, in comparison with primate CES1: (1) The helix following the primate CES1 'side door' residues (human CES1 424VMF) was extended into the 'side door' region for primate CES2 proteins (415MF for human CES2); (2) The CES1 N-terminus helix (human CES1 residues 2-14) was considerably lengthened for the primate CES2 proteins (human CES2 residues 3-23); and (3) The CES1 C-terminus helix (human CES1 residues 540-554) was significantly extended for primate CES2, including forming a second smaller predicted helix at or near the HTEL terminus for three of the four primate CES2 structures (human CES2 residues 531-549 and 554-557 respectively (Figure 1). The role of these longer helices for primate CES2 in comparison with primate CES1 awaits 3-dimensional structural analysis of primate CES2.

The phylogenetic tree reported for primate CES1 and CES2 (together with mouse CES3 and CES2 included as outgroups) (Figure 4) was calculated by the progressive alignment of 8 primate CES amino acid sequences and shows a cluster into two main groups (or families) consistent with CES1 and CES2 being products of ancestral gene duplication events. This is similar to previously published phylogenetic trees for mammalian CES [46, 57, 58] and is consistent with the results of a recent study of marsupial (opossum) and eutherian CES, which indicated that CES gene duplication events which generated ancestral mammalian CES genes predated the common ancestor for eutherian and marsupial mammals [25]. This study proposed that several ancient gene duplication events took place prior to the evolutionary appearance of mammals, generating five ancestral genes for *CES1*, *CES2, CES3*, *CES6* and *CES7*. The timing of the CES1/CES2 gene duplication event has been estimated at ~ 330 MY ago following the appearance of tetrapods during vertebrate evolution [25]. In addition to these ancestral CES gene duplication events, it is apparent that several more recent gene duplications have occurred for mammalian CES1 and CES2. For example, mouse CES1 'like' and CES2 'like' genes exist as multiple copies which are closely linked on chromosome 8 [6, 25, 26]; rat CES1 'like' and CES2 'like' genes also occur as multiple copies although in this case, they are localized on chromosomes 19 and 1, respectively [19, 49]; and a second CES1 'like' gene has been reported in human, designated as CES4, and localized on chromosome 16 within 80 kilo bases of CES1 [11, 28, 67].

 Sequence divergences observed for 29 mammalian CES gene products were used to obtain mean amino acid sequence divergence rates for primate and other eutherian mammalian CES1 and CES2 proteins, which were expressed as a percentage of amino acids substituted for each lineage per million years (MY) of primate and other mammalian evolution (Table 4). Primate CES1 sequences diverged at an average rate of 0.13% per MY, which is not significantly different to the average divergence rates observed for CES1 (0.14%/MY) and CES2 (0.16%/MY) sequences from other eutherian mammals. In contrast, primate CES2 sequences diverged at a significantly higher rate (0.24%/MY) (insert level of significance) as compared with primate CES1 and other eutherian mammalian CES1 and CES2 sequences. Studies on other gene families have shown that while most display an approximate clock-like gene divergence for family members, there are significant differences in the rate of gene divergence in a number of cases [1, 20]. The vertebrate β-globin genes have been most extensively investigated and significant variation in evolutionary rates for these genes has been demonstrated , but with an underlying molecular clock for these genes, particularly for recently duplicated genes [1]. The evolution of this gene family apparently reflects competing evolutionary events, including gene divergence as a result of the fixation of gene mutations, gene conversion events which ensure that structure-function relationships are retained, and selection pressure for divergent functions of β-globins [1, 8, 29, 70]. Mammalian CES1 and CES2 gene evolution may show a similar balance of evolutionary change resulting in differential rates of evolution for primate CES2 as compared with primate CES1.

In conclusion, the results of the present study indicate that primate CES1 and CES2 have very similar amino acid sequences with the corresponding human enzymes and share key conserved sequences and structures that have been reported for human CES1, and have family specific sequences consistent with their multimeric and monomeric subunit structures respectively. Predicted secondary and tertiary structures for baboon CES1 and CES2 showed a high degree of conservation with human CES1. Phylogeny studies using primate and other mammalian CES1 and CES2 amino acid

sequences showed that these two CES classes have undergone sequence divergence during mammalian and primate evolution, with primate CES2 showing higher amino acid substitution rates than that for primate CES1.

Acknowledgements:

Figure Legends

Figure 1: Alignment of Primate CES1 and CES2 Amino Acid Sequences

Sequence identities: * identical residues : 1 and . 2 alternative residues observed; residues involved in microsomal processing at N- and C- termini; N-glycosylation residues at 79NAT (CES1) and potential N-glycosylation sites Active site residues Ser; Glu; and His. AS: active site Side door and gate residues; Cholesterol binding Gly residue for human CES1 Z site; Disulfide bond Cys residues for human CES1 S---S; Charge clamp residues identified for human CES1 -....+; Helix (human CES1) or predicted helix; Sheet (human CES1) or predicted sheet structures

Figure 2: Predicted Tertiary Structures for Baboon CES1 and CES2

Figure 3: Gradient PAGE Zymogram of Baboon Liver and Intestine CES1 and CES2. 1: Baboon intestine CES1; 2: Baboon intestine CES2; 3: Baboon intestine CES1 and CES2; 4: Baboon liver CES1; 5: Baboon liver CES2; 6: Baboon liver CES1 and CES2. Stds: Native protein MW standards.

Figure 4: Phylogenetic tree of primate (human; chimp; baboon; and rhesus) and mouse CES sequences. Each branch of the tree is labeled with the gene name and followed by the species name. **O** designates a Catarrhine common ancestor estimated to occur at ~ 25 million years ago; **X** designates the eutherian common ancestor estimated to occur at ~ 100 million years ago. The line joining the CES1 and CES2 eutherian common ancestors designates the separation for the 2 separate lines of evolution following the proposed gene duplication event at ~ 360 million years ago.

Table Legends

Table 1: Baboon cDNA: CES1 and CES2 Sequencing Primers

Table 2: Mammalian CES1 and CES2 Genes and Proteins

*predicted sequence

Table 3: Percentage Identities for Mammalian CES1 and CES2 Sequences

h-human; c-chimpanzee; o-orangutan; b-baboon; r-rhesus

REFERENCES

1. Aguileta G, Bielawski JP, Yang Z. Evolutionary rate variation among vertebrate β globin genes: implications for dating gene family duplication events. *Gene* 2006 **380:** 21-29.
2. Ahmad S, Forgash AJ. Nonoxidative enzymes in the metabolism of insecticides. *Drug Metab Rev* 1976; **5**:141-164.
3. Becker A, Bottcher A, Lackner KJ, Fehringer P, Notka F, Aslanidis C, Schmitz G. 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.
4. Bencharit S, Edwards CC, Morton CL, Howard-Williams EL, Kuhn P, Potter PM, Redinbo MR. Multisite promiscuity in the processing of endogenous substrates by human carboxylesterase 1. *J Mol Biol*. 2006 **363**: 201-214.
5. Bencharit S, Morton CL, Xue Y, Potter PM, Redinbho MR. Structural basis of heroin and cocaine metabolism by a promiscuous human drug-processing enzyme. *Nature Struct Biol* 2003 **10**:349-356.
6. 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 **30B**:859-865.

7. Blangero J, MacCleur JW, Kammerer CM, Mott GE, Dyer TD, MacGill Jr HC. Genetic analysis of apolipoprotein A-1 in two dietary environments. *Am J Hum Genet* 1990 **47**:414-428.
8. Bromham L, Penny D. The modern molecular clock. *Nature* 2003 **4:** 216-224.
9. CES1 Gene Card GC16M054395 http://www.genecards.org/cgi-bin/carddisp.pl?gene=CES1
10. CES2 Gene Card GC16P065525 http://www.genecards.org/cgi-bin/carddisp.pl?gene=CES2
11. CES4 Gene Card GC16P054352 http://www.genecards.org/cgi-bin/carddisp.pl?gene=CES4
12. Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD. Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res* 2003 **31**:3497-3500.
13. Comuzzie AG, Cole SA, Martin L, Carey KD, Mahaney MC. The baboon as a nonhuman primate model for the study of the genetics of obesity. *Obesity Res* 2003 **11**:75-80.
14. Cygler M, Schrag JD, Sussman JL, Harel M, Silman L, Gentry MK, Doctor BP. Relationship between sequence conservation and three-dimensional structure in a large family of esterases, lipases and related proteins. *Protein Sci* 1993 **2**:366-382.
15. Diczfalusy M A, 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.
16. 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.
17. Fleming CD, Bencharit S, Edwards CC, Hyatt JL, Trurkan L, Bai B, Fraga C, Morton CL, Howard-Williams EL, Potter PM, Redinbo BM. Structural insights into drug processing by human carboxylesterase 1: tamoxifen, Mevaststin, and inhibition by Benzil. *J Mol Biol* 2005 **352**:165-177.
18. Furihata T, Hosokawa M, Nakata F, Satoh T, Chiba K. Purification, molecular cloning and functional expression of inducible acylcarnitine hydrolase in C57BL/6J mouse belonging to the carboxylesterase gene family. *Arch Biochem Biophys* 2003 **416**: 101-109.
19. Furihata T, Hosokawa M, Fujii A, Derkel M, Satoh T, Chiba K. Dexamethosone-induced methylprednisolone hemisuccinate hydrolase: its identification as a member of the rat carboxylesterase 2 family and its unique presence in plasma. *Biochem Pharm* 2005 **69**: 1287-1297.
20. Furihata T, Hosokawa M, Masuda M, Satoh T, Chiba K. Hepatocyte nuclear factor-4α plays pivotal roles in the regulation of mouse carboxylesterase 2 gene transcription in mouse liver. *Arch Biochem Biophys* 2006 **447**: 107-117.
21. Ghosh S. Cholesteryl ester hydrolase in human monocyte/macrophage: cloning, sequencing and expression of full-length cDNA. *Physiol Genomics* 2000 **2**: 1-8.
22. Gibbs PE, Witke WF, Dugaiczyk A. The molecular clock runs at different rates among closely related members of a gene family. *J Mol Evol* 1998 **46**: 552-561.
23. Goodman M, Porter CA, Czelusniak J, Page SL, Schneider H, Shoshani J, Gunnel G, Groves CP. Toward a phylogenetic classification of primates based on DNA evidence complimented by fossil evidence. *Mol Phylogenet Evol* 1998 **9**:585-598.
24. Guex N, Peitsch MC. (1997) SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modelling. *Electrophoresis* 1997 **18**:2714-2723.
25. 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* submitted for publication.
26. Hosokawa M, Furihata T, Yaginuma Y, Yamamoto N, Kayano N, Fujii A, Nagahara Y, Satoh T, Chiba K. Genomic structure and transcriptional regulation of the rat, mouse and human carboxylesterase genes. *Drug Metab Rev* 2007 39:1-15.
27. Kakenko A, Demop K, Onoe T. Heterogeneity of esterases and cell types in rat liver. *Biochim Biophys Acta* 1972 **284**:128-135.
28. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Hussler D. The human genome browser at UCSC. *Genome Res* 2002 **12**:994-1006.
29. Kimura M. *The Neutral Theory of Evolution*. Cambridge University Press 1983
30. Kopp J, Schwede T. The SWISS-MODEL Repository of annotated three-dimensional protein structure homology models. *Nucleic Acids Res* 2004 **32**:D230-D234.
31. 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. *Biochem* 1993 **32**:11606-11617.

32. Langmann T, Becker A, Aslanidis C, Notka F, Ullrich H, Schwer H, Schmitz G. Structural organization and characterization of the promoter region of a human carboxylesterase gene. *Biochim Biophys Acta* 1997 **1350**: 65-74.
33. Leinweber FJ. Possible physiological roles of carboxyl ester hydrolases. *Drug Metab Rev* 1987 **18**:379-439.
34. 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.
35. MacCleur, JW, Kammerer CM, VandeBerg JL, Cheng M-L, Mott GE, McGill Jnr HC. Detecting genetic effects on lipoprotein phenotypes in baboon: a review of methods and preliminary findings. *Genetica* 1987 **73**:159-168.
36. Marsh S, Xiao M, Yu J, Ahluwalia R, Minton M, Freimuth RR, Kwok P-Y, McLeod HL. Pharmacogenomic assessment of carboxylesterases 1 and 2. *Genomics* 2004 **84**:661-668.
37. McBrearty S, Jablonski NG. First fossil chimpanzee. *Nature* 2005 **437**:105-108.
38. McGuffin LJ, Bryson K, Jones DT. The PSIPRED protein structure prediction server. *Bioinformatics* 2000 **16**: 404-405.
39. Miyazaki M, Kamiie K, Soeta S, Taira H, Yamashita T. Molecular cloning and characterization of a novel carboxylesterase-like protein that is physiologically present at high concentrations in the urine of domestic cats (Felis Catus). *Biochem J* 2003 **370**:101-110.
40. Morgan D, Grant KA, Gage HD, Mach RH, Kaplan JR, Prioleau O, Nader SH, Buchheimer N, Ehrenkaufer RL, Nader MA. Social dominance in monkeys: dopamine D2 receptors and cocaine self-administration. Nature *Neuroscience* 2002 **5**:169-174.
41. Mouse Genome Browser. 2007 http://mouse.perlegen.com/cgi-bin/gbrowse/mouse_b04/?name=CES3
42. Munro S, Pelham HR. A C-terminal signal prevents secretion of luminal ER proteins. *Cell* 1987 **48**:899-907.
43. Murphy WJ, Eizirik E, Johnson, W.E, Zhang YP, Ryder OA, O'Brien SJ. (2001) Molecular phylogenetics and the origins of placental mammals. *Nature* 2001 **409**:614–618.
44. Ozols J. Isolation, properties, and the complete amino acid sequence of a second form of 60-kDa glycoprotein esterase. Orientation of the 60-kDa proteins in the microsomal membrane. *J Biol Chem* 1989 **264**:12533–12545.
45. Peitsch MC. Protein modeling by E-mail. *Bio Technology* 1995 **13**:658-660.
46. Pindel EV, Kedishvili NY, Abraham TL, Brzezinski MR, Zhang J, Dean RA, Bosron WF. 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.
47. Potter PM, Wolverton JS, Morton CL, Wierdl M, Danks M. 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.
48. Raaum R, Sterner KN, Noviello CM, Stewart C-B, Disotell TR. Catarrhine primate divergence dates estimated from complete mitochondrial genomes: concordance with fossil and nuclear DNA evidence. *J Hum Evol* 2005 **48:** 237-257.
49. Rat Genome Database. 2007 http://rgd.mcw.edu/tools/genes/genes_view.cgi?id=621508
50. Redinbo MR, Potter PN. Mammalian carboxylesterases: from drug targets to protein therapeutics. *Drug Discov Today* 2005 **10**: 313-320.
51. Rhesus macaque genome sequencing and analysis consortium. Evolutionary and biomedical insights from the rhesus macaque genome. *Science* **316**:222-234.
52. Ricaurte GA, Yuan J, Hatzidimitriou G, Cord BJ, McCann UD. Severe dopaminergic neurotoxicity in primates after a common recreational dose regimen of MDMA ("Ecstasy"). *Science* 2002 **297**:2260 – 2263.
53. Robbi M, Beaufay H. The COOH terminus of several liver carboxylesterases targets these enzymes to the lumen of the endoplasmic reticulum. *J Biol Chem* 1991 **266**:20498-20503.
54. Rogers J, Hixson JE. Baboons as an animal model for genetic studies of common human disease. Am J Human *Genetics* 1997 **61**:489-493.
55. Sanghani SP, Quinney SK, Fredenberg TB, Davis WI, Murry DJ, Bosron WF. 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.
56. Satoh T, Hosokawa M. The mammalian carboxylesterases: from molecules to functions. *Ann Rev Pharmacol Toxicol* 1998 **38**:257-288.
57. Satoh T, Hosokawa M. Structure, function and regulation of carboxylesterases. *Chem-Biol Interactions* 2006 **162**:195-211.
58. Satoh H, Taylor P, Bosron WF, Sanghani SP, Hosokawa M, La Du BN. Current progress on esterases: from molecular structure to function. *Drug Metab Dispos* 2002 **30**:488-493.

59. Schewer H, Langmann T, Daig R, Becker A, Aslandis C, Schmitz G. Molecular cloning and characterization of a novel putative carboxylesterase, present in human intestine and liver. *Biochem Biophys Res Commun* 1997 **233**:117-120.
60. Schwede T, Kopp J, Guex N, Peitsch MC. SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Res* 2003 **31**:3381-3385.
61. Shibita F, Takagi Y, Kitajima M, Kuroda T, Omura T. Molecular cloning and characterization of a human carboxylesterase gene. *Genomics* 1993 **17**:76-82.
62. Tsujita T, Okuda H. Palmitoyl-coenzyme A hydrolyzing activity in rat kidney and its relationship with carboxylesterase. *J Lipid Res* 1993 **34:**1773-1781.
63. von Heijne G. Patterns of amino acids near signal-sequence cleavage sites. *Eur J Biochem* 1983 **133**:17-21.
64. Wambutt R, Huebner D, Mewes HW, Amid C, Osanger a, Fobo G, Han R, Wiemann S. The German cDNA Consortium. Submitted to *EMBL/Gen/DDBJ Databases* 2004.
65. 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**: 33218-26.
66. Woodburne MO, Rich TH, Springer MS. The evolution of tribospheny and the antiquity of mammalian clades. *Mol Phylogenet Evol* 2003 **28**:360-385.
67. Yan B, Matoney L, Yang D. Human carboxylesterases in term placentae: enzymatic characterization, molecular cloning and evidence for existence for multiple forms. *Placenta* 1999 **20**:599-607.
68. Yang Z, Rannala B. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Mol Biol Evol* 2005 **23**:212-226.
69. 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.
70. Zuckerkandl E. On the molecular evolutionary clock. *J Mol Evol* 1987 **26**: 34-4
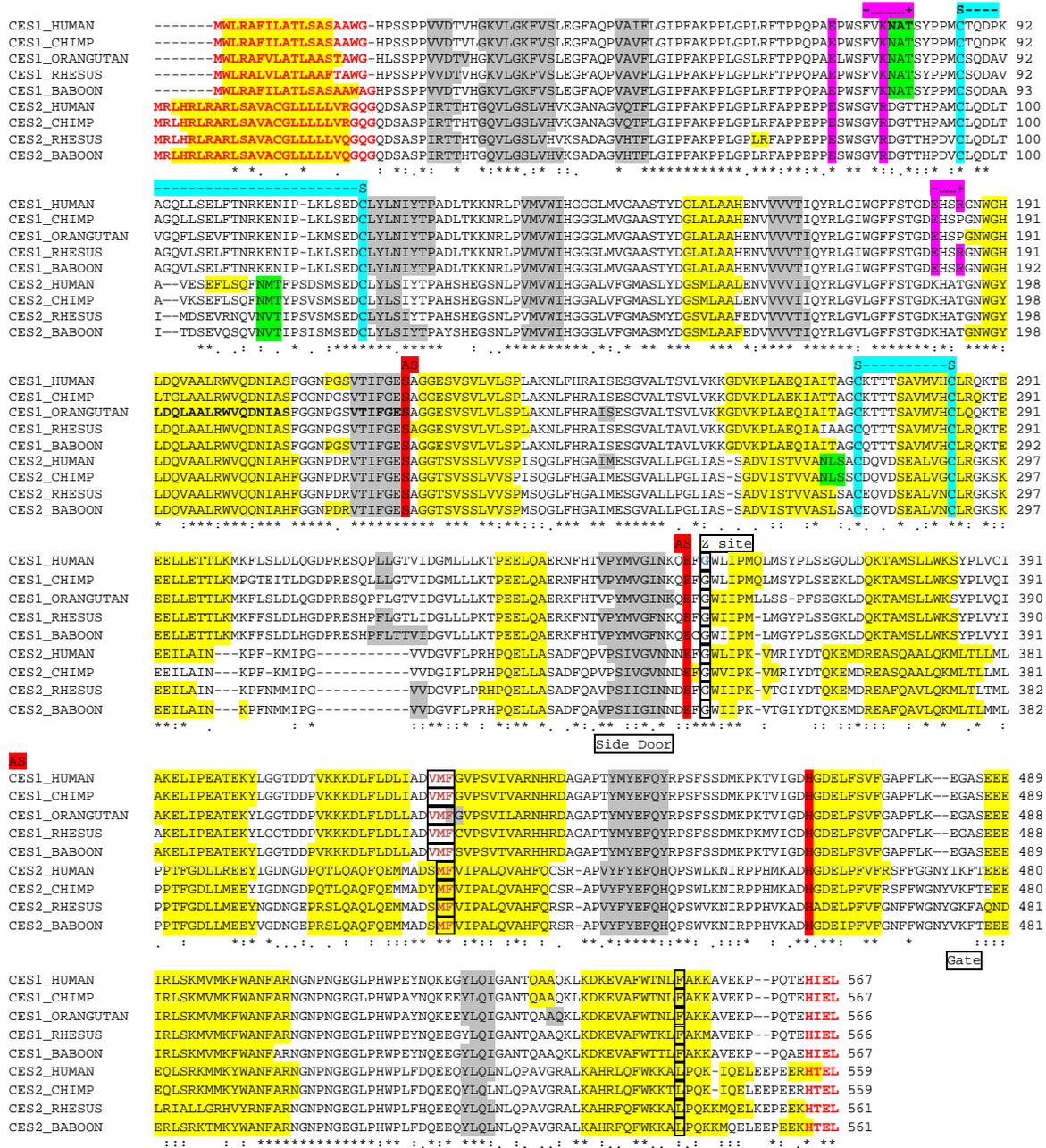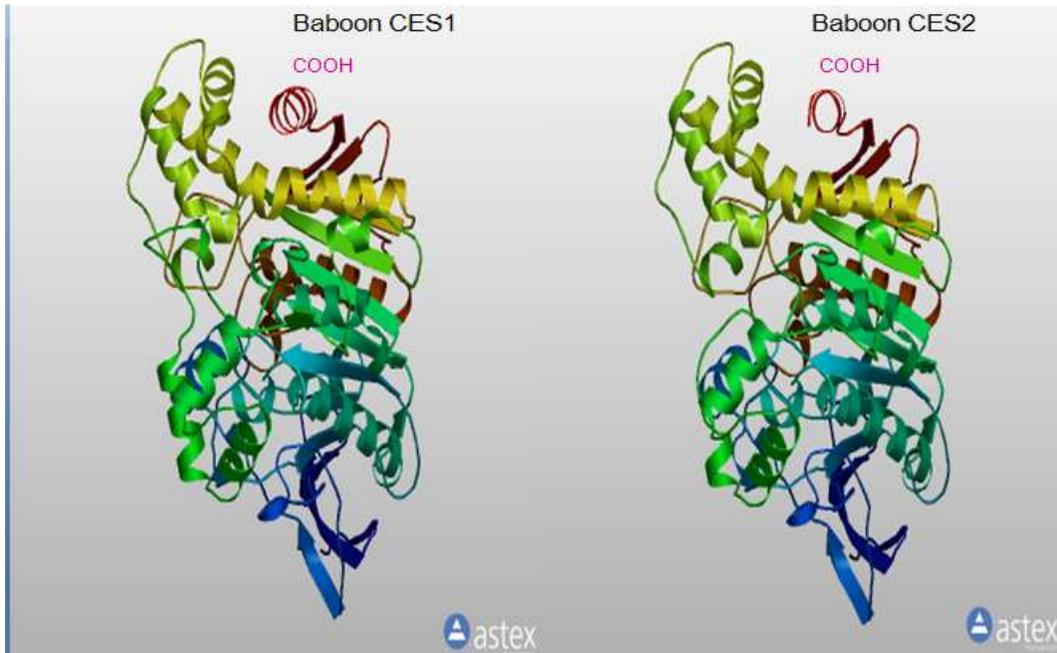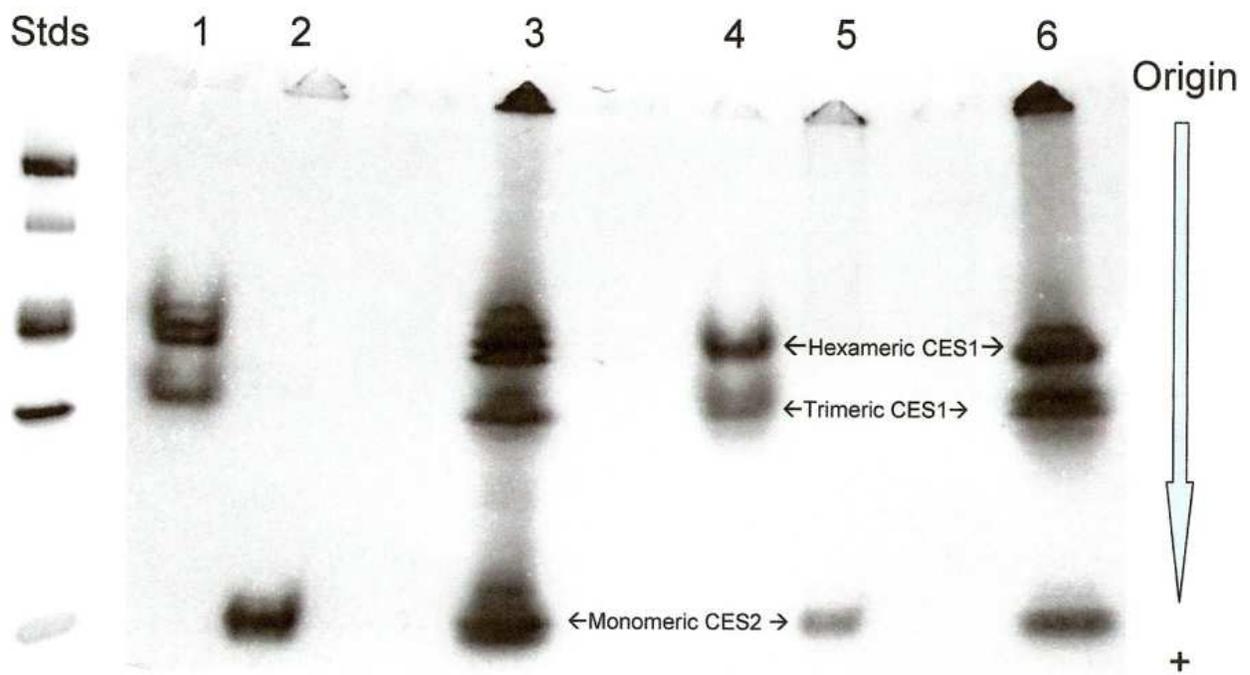
Figure 1:

```
CES1_HUMAN     -------MWLRAFILATLSASAAWG-HPSSPPVVDTVHGKVLGKFVSLEGFAQPVAIFLGIPFAKPPLGPLRFTPPQPAEPWSFVKNATSYPPMCTQDPK 92
CES1_CHIMP     -------MWLRAFILATLSASAAWG-HPSSPPVVDTVLGKVLGKFVSLEGFAQPVAVFLGIPFAKPPLGPLRFTPPQPAEPWSFVKNATSYPPMCTQDPK 92
CES1_ORANGUTAN -------MWLRAFVLATLAASTAWG-HLSSPPVDTVHGKVLGKFVSLEGFAQPVAVFLGIPFAKPPLGSLRFTPPQPAEPWSFVKNATSYPPMCSQDAV 92
CES1_RHESUS    -------MWLRALVLATLAAFTAWG-HPSSPPVVDTVHGKVLGKFVSLEGFAQPVAVFLGIPFAKPPLGPLRFTPPQPAEPWSFVKNATSYPPMCSQDAV 92
CES1_BABOON    -------MWLRAFILATLSASAAWAGHPSSPPVVDTVHGKVLGKFVSLEGFAQPVAVFLGIPFAKPPLGPLRFTPPQPAEPWSFVKNATSYPPMCSQDAA 93
CES2_HUMAN     MRLHRLRARLSAVACGLLLLLVRGQGQDSASPIRTTHTGQVLGSLVHVKGANAGVQTFLGIPFAKPPLGPLRFAPPEPPSWSGVRDGTTHPAMCLQDLT 100
CES2_CHIMP     MRLHRLRARLSAVACGLLLLLVRGQGQDSASPIRTTHTGQVLGSLVHVKGANAGVQTFLGIPFAKPPLGPLRFAPPEPPSWSGVRDGTTHPAMCLQDLT 100
CES2_RHESUS    MRLHRLRARLSAVACGLLLLLVQGQGQDSASPIRTTHTGQVLGSLVHVKSADAGVHTFLGIPFAKPPLGPLRFAPPEPPSWSGVRDGTTHPDVCLQDLT 100
CES2_BABOON    MRLHRLRARLSAVACGLLLLLVQGQGQDSASPIRTTHTGQVLGSLVHVKSADAGVHTFLGIPFAKPPLGPLRFAPPEPPSWSGVRDGTTHPDVCLQDLT 100
                       *  *. .*  .* :  *:.*:  *  *:***.:* ::.    *  *********** .***:**:*.* ** *::.*::* :* **

CES1_HUMAN     AGQLLSELFTNRKENIP-LKLSEDCYLNIYTPADLTKKNRLPVMVWIHGGGLMVGAASTYDGLALAAHENVVVVTIQYRLGIWGFFSTGDEHSRGNWGH 191
CES1_CHIMP     AGQLLSELFTNRKENIP-LKLSEDCYLNIYTPADLTKKNRLPVMVWIHGGGLMVGAASTYDGLALAAHENVVVVTIQYRLGIWGFFSTGDEHSPGNWGH 191
CES1_ORANGUTAN VGQFLSEVFTNRKENIP-LKMSEDCYLNIYTPADLTKKNRLPVMVWIHGGGLMVGAASTYDGLALAAHENVVVVTIQYRLGIWGFFSTGDEHSPGNWGH 191
CES1_RHESUS    AGQVLSELFTNRKENIP-LKLSEDCYLNIYTPADLTKKNRLPVMVWIHGGGLMVGAASTYDGLALAAHENVVVVTIQYRLGIWGFFSTGDEHSRGNWGH 191
CES1_BABOON    AGQVLSELFTNRKENIP-LKLSEDCYLNIYTPADLTKKNRLPVMVWIHGGGLMVGAASTYDGLALAAHENVVVVTIQYRLGIWGFFSTGDEHSRGNWGH 192
CES2_HUMAN     A--VESEFLSQFNMTFPSDSMSEDCLYLSIYTPAHSHEGSNLPVMVWIHGGALVFGMASLYDGSMLAALENVVVVTIQYRLGVLGFFSTGDKHATGNWGY 198
CES2_CHIMP     A--VKSEFLSQFNMTYPSVSMSEDCLYLSIYTPAHSHEGSNLPVMVWIHGGALVFGMASLYDGSMLAALENVVVVTIQYRLGVLGFFSTGDKHATGNWGY 198
CES2_RHESUS    I--MDSEVRNQVNVTIPSVSMSEDCLYLSIYTPAHSHEGSNLPVMVWIHGGGLVFGMASMYDGSVLAAFEDVVVVTIQYRLGVLGFFSTGDKHATGNWGY 198
CES2_BABOON    I--TDSEVQSQVNVTIPSISMSEDCLYLSIYTPAYSHEGSNLPVMVWIHGGGLVFGMASMYDGSMLAAFEDVVVVTIQYRLGVLGFFSTGDKHATGNWGY 198
                         **. :: . * :*****.*****  :: ..*********.*:.* ** *** *:*** ****** *******:* ****:

CES1_HUMAN     LDQVAALRWVQDNIASFGGNPGSVTIFGESAGGESVSVLVLSPLAKNLFHRAISESGVALTSVLVKKGDVKPLAEQIAITAGCKTTTSAVMVHCLRQKTE 291
CES1_CHIMP     LTGLAALRWVQDNIASFGGNPGSVTIFGESAGGESVSVLVLSPLAKNLFHRAISESGVALTSVLVKKGDVKPLAEKIATTAGCKTTTSAVMVHCLRQKTE 291
CES1_ORANGUTAN LDQLAALRWVQDNIASFGGNPGSVTIFGESAGGESVSVLVLSPLAKNLFHRAISESGVALTSVLVKKGDVKPLAEQIAITAGCKTTTSAVMVHCLQQKTE 291
CES1_RHESUS    LDQLAALHWVQDNIASFGGNPGSVTIFGESAGGESVSVLVLSPLAKNLFHRAISESGVALTAVLVKKGDVKPLAEQIAIAAGCKTTTSAVMVHCLRQKTE 291
CES1_BABOON    LDQLAALRWVQDNIASFGGNPGSVTIFGESAGGESVSVLVLSPLAKNLFHRAISESGVALTAVLVKKGDVKPLAEQIAITAGCQTTTSAVMVHCLRQKTE 292
CES2_HUMAN     LDQVAALRWVQQNIAHFGGNPDRVTIFGESAGGTSVSSLVVSPISQGLFHGAIMESGVALLPGLIAS-SADVISTVVANLSACDQVDSEALVGCLRGSK 297
CES2_CHIMP     LDQVAALRWVQQNIAHFGGNPDRVTIFGESAGGTSVSSLVVSPISQGLFHGAIMESGVALLPGLIAS-SGDVISTVVANLSSCDQVDSEALVGCLRGSK 297
CES2_RHESUS    LDQVAALRWVQQNIAHFGGNPDRVTIFGESAGGTSVSSLVVSPMSQGLFHGAIMESGVALLPGLIAS-SADVISTVVASLSACEQVDSEALVNCLRGSK 297
CES2_BABOON    LDQVAALRWVQQNIAHFGGNPDRVTIFGESAGGTSVSSLVVSPMSQGLFHGAIMESGVALLPGLIAS-SADVISTVVASLSACEQVDSEALVNCLRGSK 297
                *  :***:***:*** *****. *** **:**::.*** ** ****** :  . :: .* :.*.  * .:* **: :**

CES1_HUMAN     EELLETTLKMKFLSLDLQGDPRESQPLLGTVIDGMLLLKTPEELQAERNFHTVPYMVGINKQEFGWLIPMQLMSYPLSEGQLDQKTAMSLLWKSYPLVCI 391
CES1_CHIMP     EELLETTLKMPGTEITLDGDPRESQLLLGTVIDGMLLLKTPEELQAERNFHTVPYMVGINKQEFGWLIPMQLMSYPLSEEKLDQKTAMSLLWKSYPLVQI 391
CES1_ORANGUTAN EELLETTLKMKFLSLDLQGDPRESQPFLGTVIDGVLLLKTPEELQAERKFHTVPYMVGINKQEFGWIIPMLLSS-PFSEGKLDQKTAMSLLWKSYPLVQI 390
CES1_RHESUS    EELLETTLKMKFFSLDLHGDPRESHPFLGTLIDGLLLLKTPEELQAERNFHTVPYMVGFNKQEFGWIIPM-LMGYPLSEGKLDQKTAMSLLWKSYPLVYI 390
CES1_BABOON    EELLETTLKMKFFSLDLHGDPRESHPFLTTVIDGVLLLKTPEELQAERKFHTVPYMVGFNKQECGWIIPM-LMGYPLSEGKLDQKTAMSLLWKSYPLVYI 391
CES2_HUMAN     EEILAIN---KPF-KMIPG-----------VVDGVFLPRHPQELLASADFQPVSIVGVNNNEFGWLIPK-VMRIYDTQKEMDREASQAALQKMLTLLML 381
CES2_CHIMP     EEILAIN---KPF-KMIPG-----------VVDGIFLPRHPQELLASADFQPVSIVGVNNDEFGWLIPK-VMRIYDTQKEMDREASQAALQKMLTLLML 381
CES2_RHESUS    EEILAIN---KPFNMMIPG-----------VVDGVFLRHPQELLASADFQAVPSIIGINNDEFGWIIPK-VTGIYDTQKEMDREAFQAVLQKMLTLTML 382
CES2_BABOON    EEILAIN---KPFNMMIPG-----------VVDGVFLPRHPQELLASADFQAVPSIIGINNDEFGWIIPK-VTGIYDTQKEMDREAFQAVLQKMLTMML 382
                **:*  .        : *         ::*:**::* *:** *.  *.  *:* .:**:*** .     :: ::*::*.   * *   . *  :

CES1_HUMAN     AKELIPEATEKYLGGTDDTVKKKDLFLDLIADVMFGVPSVIVARNHRDAGAPTYMYEFQYRPSFSSDMKPKTVIGDRGDELFSVFGAPFLK--EGASEEE 489
CES1_CHIMP     AKELIPEATEKYLGGTDDPVKKKDLFLDLIADVMFGVPSVTVARNHRDAGAPTYMYEFQYRPSFSSDMKPKTVIGDRGDELFSVFGAPFLK--EGASEEE 489
CES1_ORANGUTAN AKELIPEATEKYLGGTDDPVKKKDLFLDLLADVMFGVPSVILARNHRDAGAPTYMYEFQYRPSFSSDMKPKTVIGDRGDELFSVFGAPFLK--EGASEEE 488
CES1_RHESUS    AKELIPEAIEKYLGGTDDPVKKKDLFLDLIADVMFCVPSVIVARHHRDAGAPTYMYEFQYRPSFSSDMKPKMVIGDRGDELFSVFGAPFLK--EGASEEE 488
CES1_BABOON    AKELIPEATEKYLGGTDDPVKKKDLFLDLLADVMFGVPSVIVARHHRDAGAPTYMYEFQYRPSFSSDMKPKTVIGDRGDELFSVFGAPFLK--EGASEEE 489
CES2_HUMAN     PPTFGDLLREEYIGDNGDPQTLQAQFQEMMADSMFVIPALQVAHFQCSR-APVYFEFQHQPSWLKNIRPPHMKADRGDELPFVFRSFFGGNYIKFTEEE 480
CES2_CHIMP     PPTFGDLLMEEYIGDNGDPQTLQAQFQEMMADYMFVIPALQVAHFQCSR-APVYFEFQHQPSWLKNIRPPHMKADRGDELPFVFRSFFWGNYVKFTEEE 480
CES2_RHESUS    PPTFGDLLMEEYNGDNGEPRSLQAQLQEMMADSMFVIPALQVAHFQRSR-APVYFEFQHQPSWVKNIRPPHVKADHADELPFVFGNFFWGNYGKFAQND 481
CES2_BABOON    PPTFGDLLMEEYNGDNGEPRSLQAQLQEMMADSMFVIPALQVAHFQRSR-APVYFEFQHQPSWVKNIRPPHVKADRGDELPFVFGNFFWGNYVKFTEEE 481
                  .   *:*  *** ...:. .: :**.** **.:*::.* :* :. .  **.*:****:**: .::::* :  :: .**:**: **  *     :::::

CES1_HUMAN     IRLSKMVMKFWANFARNGNPNGEGLPHWPEYNQKEGYLQIGANTQAAQKLKDKEVAFWTNLFAKKAVEKP--PQTEHIEL 567
CES1_CHIMP     IRLSKMVMKFWANFARNGNPNGEGLPHWPAYNQKEEYLQIGANTQAAQKLKDKEVAFWTNLFAKKAVEKP--PQTEHIEL 567
CES1_ORANGUTAN IRLSKMVMKFWANFARNGNPNGEGLPHWPAYNQKEEYLQIGANTQAAQKLKDKEVAFWTNLFAKKAVEKP--PQTEHIEL 566
CES1_RHESUS    IRLSKMVMKFWANFARNGNPNGEGLPHWPEYNQEEGYLQIGANTQAAQKLKDKEVAFWTKLFAKMAVEKP--PQTEHIEL 566
CES1_BABOON    IRLSKMVMKFWANFARNGNPNGEGLPRWPEYNQEEGYLQIGANTQAAQKLKDKEVAFWTTLFAKKAVEKP--PQAEHIEL 567
CES2_HUMAN     EQLSRKMMKYWANFARNGNPNGEGLPHWPLFDQEEYLQLNLQPAVGRALKAHRLQFWKKALPQK-IQELEEPEERHTEL 559
CES2_CHIMP     EQLSRKMMKYWANFARNGNPNGEGLPHWPLFDQEEYLQLNLQPAVGRALKAHRLQFWKKTLPQK-IQELEEPEERHTEL 559
CES2_RHESUS    LRIALLGRHVYRNFARNGNPNGEGLPHWPLFHQEEYLQLNLQPAVGRALKAHRFQFWKKALPQKKMQELKEPEEKHTEL 561
CES2_BABOON    ERLSRKTMKYWANFARNGNPNGEGLPHWPLFDQEEYLQLNLQPAVGRALKAHRFQFWKKALPQKKMQELEEPEEKHTEL 561
                :::    : :  *************:** :.:.*  ***:  :.  ..: ** :.  **.. :.:  :::   *: .* **
```

Labels within figure: S----, S, AS, Z site, Side Door, Gate

Figure 2:

Figure 3:

Figure 4: