Molecular cloning and cDNA characterization of *Camelus dromedarius* putative cytochrome P450s 1A, 2C, and 3A

H.M. Saeed[1,2], M.S. Alanazi[1], M.A. Shalaby[3], O. Alshahrani[1], F.S. Ataya[4], A.A. Pathan[1] and Z.A Abduljaleel[1]

[1]Department of Biochemistry, College of Science, King Saud University, Riyadh, Saudi Arabia
[2]Department of Bioscience and Technology, Institute of Graduate Studies and Research, Alexandria University, Alexandria, Egypt
[3]Genetic Engineering and Biotechnology Research Institute, City for Scientific Research and Technology Applications, Alexandria, Egypt
[4]Department of Molecular Biology, Genetic Engineering Division, National Research Center, Dokki, Cairo, Egypt

Corresponding author: H.M. Saeed
E-mail: hesham25166@hotmail.com / hsaeed1@KSU.EDU.SA

ABSTRACT. The domesticated one-humped Arabian camel, *Camelus dromedarius*, is one of the most important animals in the Arabian Peninsula. Most of its life, this animal is exposed to both intrinsic and extrinsic genotoxic factors that are known to cause gross metabolic alterations in many organisms. This study determined the full length coding sequence of 3 cytochrome P450s cDNAs; namely, CYP450 1A1, CYP450 2C and CYP450 3A using reverse transcription polymerase chain reaction. The *C. dromedarius* CYP450s 1A1, 2C, and 3A have open reading frames of 1563, 1473, and 1566 bp and cDNAs that encode proteins of 520, 490, and 521 amino acid residues, respectively. The molecular weights

calculated for CYP1A1, 2C, and 3A were found to be 58.651, 56.03, and 58.594 kDa, while the predicted calculated isoelectric points using a computer algorithm were 7.315, 6.579, and 9.46. The deduced amino acid sequences of these CYPs showed the membrane anchored signal peptide, the conserved proline-rich amino terminus and the characteristic heme-binding signature localized near the carboxy terminus of the protein.

Key words: Arabian camel; *Camelus dromedarius*; Cytochrome P450; Xenobiotics; Reverse transcription polymerase chain reaction; Mixed-functions oxidases

INTRODUCTION

The Arabian one-humped camel (*Camelus dromedarius*) belongs to the family Camelidae, which has 6 camel-like animals (1-hump and 2-hump camels, llamas, alpaca, guanaco, and vicuna). This camel is found in the Arabian deserts and arid and semiarid areas of the Middle East (Thomas et al., 1985; Hussein et al., 1992; Al-Khedhairy, 2004). The Arabian camel has unique physiological modifications and is able to live in harsh desert conditions, such as high temperature, direct exposure to sunlight and lack of water resources, with this animal being able to survive without drinking water for weeks (Hussein et al., 1992).

While there has been an increasing number of studies on camels in recent years, pharmacological and biomolecular research of camels remains limited. The Arabian camel genome consists of 74 chromosomes. The DNA sequence of camels exhibits wide variations with other species, including human; however, functional similarities exist. Therefore, studies of camel protein represent a potentially viable target for human applications (Al-Khedhairy, 2004).

Like other living organisms, the Arabian camel is continuously exposed to intrinsic and extrinsic agents that, if not treated properly, may result in mutation and cell death. Many enzymes are involved in the clearance of such compounds such as Phase I and Phase II drug metabolizing enzymes. The clearance of xenobiotics involves both activation (Phase I) and detoxification (Phase II) reactions. Cytochrome P450 monooxygenases (CYPs or P450) are heme-containing proteins belonging to Phase I drug metabolizing enzymes. These enzymes use molecular oxygen and the hydride donor, NADPH, to effect the overall oxidative insertion of one oxygen atom into the substrate (Jung et al., 2011; Nelson, 2011). Cytochrome P450s are involved in the metabolism of a broad range of substrates, and catalyze a variety of chemical reactions, including the conversion of both endogenous and exogenous compounds to more soluble hydrophilic metabolites that are easily removed from the body (Nelson and Strobel, 1987; Nelson et al., 1993). Cytochrome P450s are mixed-functions oxidases that represent one of the body's most important defenses against chemical-induced toxicity. However, in some situations, the metabolism of some chemicals by P450s is undesirable, and may lead to toxic or reactive intermediates, resulting in target organ toxicity and/or carcinogenic insult (Rooney et al., 2004; Nebert and Dalton, 2006; Stiborova et al., 2011).

Most CYP enzymes are expressed constitutively in a variety of mammalian tissues, such as liver, kidney, lung, adrenal gland, and gonads (Seliskar and Rozman, 2007). Many other CYPs are inducible enzymes, and their expression levels increase markedly when exposed to various chemicals, such as ethanol and polychlorinated dibenzo-ρ-dioxins (Uppstad et al., 2010). P450 proteins are conveniently arranged into families and subfamilies, based

on the percentage of amino acid sequence identity. For instance, enzymes that share ≥40% identity are assigned to a particular family designated by an Arabic numeral, whereas those sharing ≥55% identity are assigned to a particular subfamily designated by a letter (Hasler et al., 1999). With the advent of sequencing techniques and the sequencing of genomes of many different organisms, the number of sequenced P450s is rising. At present, the P450 superfamily consists of 12,456 named sequences, with about 6000 more known but not yet named.

The human genome contains 57 genes encoding P450 enzymes and 58 pseudogenes that are distributed in 18 P450 families. Seven of these genes are Type I enzymes, which are found in the mitochondria, while the other 50 genes encode Type II enzymes, which are found in the endoplasmic reticulum (Guengerich, 2004). One-fourth of the human P450s is not well characterized and are termed orphans (Nelson, 2011). The Arabian camel exhibits P450-dependent activities in hepatic and extrahepatic tissues, such as the kidney, lung, spleen, tongue, and the hump (el Sheikh et al., 1991; Raza et al., 1998, 2004; Jung et al., 2011). The present study aimed to isolate and characterize the full-length coding sequence cDNAs of *C. dromedarius* cytochrome P450 1A, 2C and 3A and to assess the degree of similarity of the deduced proteins with those of other ungulates.

MATERIAL AND METHODS

Chemicals

All of the chemicals used in this study were of analytical reagent, molecular biology, or chromatographic grade, as appropriate. Water was de-ionized and distilled.

Animals and tissue preparation

Liver samples from 3 adult male Arabian one-humped camels (*C. dromedarius*) were obtained from the local slaughterhouse, after the animals were killed under the observation of a skilled veterinarian. The tissue samples selected for RNA preparation and analysis were immediately submerged in RNAlater® solution (Qiagen, Ambion, Courtabeuf, France) to avoid RNA degradation. The samples were stored at 4ºC for 24 h, and then stored at 20ºC until use.

RNA isolation and cDNA synthesis

Samples of 30-60 mg of the preserved tissues were homogenized in RLT lysis buffer (Qiagen) supplemented with 1% 2-mercaptoethanol, using a rotor-stator homogenizer (MEDIC TOOLS, Switzerland). Total RNA was extracted using the RNeasy Mini Kit (Qiagen, Germany), with a DNase digestion step, following manufacturer protocols. Elution was performed with 50 μL nuclease-free water. Concentration, purity, and quality of the isolated RNA were determined by the Agilent 2100 Bioanalyzer System and Agilent Small RNA analysis kit, following manufacturer protocols (Agilent Technologies, Waldbronn, Germany). Total RNA in aliquots of 2 μg was retro-transcribed into single-stranded cDNA by the ImProm-II Reverse Transcription System (A3800, Promega, Madison, USA). Complementary DNA was synthesized by reverse transcription, and used as a template for the amplification of the camel CYPs being studied.

PCR

PCR was carried out in a final volume of 50 µL, containing 25 µL 2X high-fidelity PCR master mix (GE Healthcare, USA), 5 µL c-DNA , 3 µL 30 pM of each primer, CYP1A forward: 5'- *GGATCC*ATGTTCTCTGTGTTTGGACTCTCC-3' and reverse: 5'- *GCGGCC*GCCTAAG AGCGCATATGCACCT-3'; CYP3A forward: 5'- ATGGACCTGATCCCAAGCTTTTCCTT-3' and reverse: 5'-GTCCTTAGGGAAAGTCAGGCTTCATT-3'; and CYP2C forward: 5'- ATGGATCTCTTCATAGTCCTGGTG-3' and reverse: 5'- GACTGGGAATGAAACG-GACCTCATAGA-3'. The PCR condition was 1 cycle at 95°C for 5 min, followed by 30 cycles at 94°C for 1 min, 55°C for 1 min, and 72°C for 2 min. The final extension step was carried out at 72°C for 5 min. The PCR products were analyzed on 1% agarose gel stained with ethidium bromide, at a concentration of 0.5 µg/mL.

Cloning and sequencing of the PCR products

To ligate the PCR products generated onto pGEM®-T Easy vector (Promega Co. Cat No. A1360), 2 µL of the product was placed in a clean 0.5-mL tube, to which 1 µL pGEM®-T-Easy vector (50 ng) and 1 µL 10X ligase buffer were added, followed by the addition of 2 U ligase enzyme. The final volume of the ligation reaction was adjusted to 10 µL, by the addition of nuclease-free water. The tube was incubated at 16°C for 16 h. Transformation of *Escherichia coli* JM109 competent cells was carried out according to Sambrook et al. (1989). The recombinant *E. coli* JM109 harboring the pGEM®-T-Easy vector was screened on selective LB/IPTG/X-gal/Ampicillin/agar plates. Plasmids were prepared from positive clones by the PureYield Plasmid Miniprep System (Cat No. A1222, Promega, Madison, WI, USA).

Sequencing of the PCR products cloned onto pGEM®-T-Easy vector (4 different clones) was carried out according to Sanger et al. (1977) by the MegaBACE 1000 DNA Sequencing System (Pharmacia/Amersham Co., USA). The chain termination sequencing reaction was conducted by the DYEnamic ET terminator kit, as an integral part of the MegaBACE 1000 DNA sequencing system. The sequencing reaction products were purified by the DyeEx 2.0 Spin Kit (Qiagen, Cat No. 63206), and applied to the MegaBace 1000 Sequencing machine. The cDNA and amino acid sequences of the camel CYP1A, CYP3A, and CYP2C were aligned with those of other mammals using the ClustalW and DNAStar programs.

Cloning and expression of *C. dromedarius* CYP1A into the pET28a(+) vector

pGEM®-T-Easy plasmid that carries the *C. dromedarius* CYP1A cDNA insert was digested with *Bam*HI and *Not*I restriction enzymes to release the cloned insert. After agarose gel electrophoresis, the CYP1A fragment was purified from the agarose gel and subcloned into the pET-28a(+) expression vector. The pET-28a(+) (Novagen Co.) plasmid carries an N-terminal His-Tag/thrombin/T7 configuration. The expression of the cloned CYP1A cDNA in pET28a(+) is under the control of the T7 promoter. The pET-28a(+) was digested with *Bam*HI and *Not*I restriction enzymes, and treated with shrimp alkaline phosphatase (Promega), according to the method of Sambrook et al. (1989). Ligation was carried out in a 10-µL final volume that contained 2 µL pET-28a(+) plasmid (20 ng), 2 µL CYP1A cDNA insert (20 ng), 1 µL 10X ligase buffer, and 1 µL 2 U ligase enzyme. The final volume of the ligation reaction was adjusted to 10 µL by the addition of

nuclease-free distilled water. The tube was incubated at 16°C overnight. Transformation of *E. coli* BL21(DE3) pLysS (Promega) competent cells was carried out according to the method of Sambrook et al. (1989). The recombinant *E. coli* BL21(DE3) pLysS harboring the pET-28a(+) vector was screened on selective LB/IPTG/X-gal/Kanamycin/chloramphenicol/agar plates, and by using the colony PCR strategy utilizing CYP1A gene specific primers. Plasmids were prepared from positive clones using the PureYield Plasmid Miniprep System (Promega, Cat No. A1222).

Expression of recombinant *C. dromedarius* CYP1A in *E. coli* BL21(DE3) pLysS

A 1-L Erlenmeyer flask containing 200 mL of terrific broth supplemented with kanamycin and chloramphenicol (Sigma/Aldrich Co., USA) at concentrations of 50 and 35 μg/mL, respectively, were inoculated with 2 mL overnight culture of *E. coli* BL21(DE3) pLysS pET-28a(+) carrying the CYP1A gene. The flasks were shaken at 250 rpm at 37°C until the absorbance at 600 nm was 0.2, at which point 0.5 mM δ-aminolevulinic acid was added to the culture medium. The incubation was continued until the absorbance at 600 nm reached 0.8. Then, 1 mM IPTG was added to the culture and the temperature was adjusted to 30°C and 150 rpm for 20 h. After the incubation period, the cells were harvested by centrifugation at 8000 rpm at 4°C for 20 min, washed with 50 mM potassium phosphate buffer (pH 7.5, containing 5.0 mM $MgCl_2$), and pelleted for a second time. The cells were then resuspended in 10 mL 50 mM potassium phosphate buffer (pH 7.5, containing 5.0 mM $MgCl_2$, 0.2 mM DTT, and 10% glycerol), and sonicated using 4X 15 s pulses. After the first pulse, 1 mM PMSF, 0.1 μg/mL leupeptin, and 0.04 U/mL aprotinin were added to the tubes and then sonication was continued. Cells debris was removed by centrifugation at 12,000 rpm at 4°C for 10 min after which the supernatant was collected for protein assays and Western blotting.

Total protein determination

The total protein concentration was assayed by the method of Bradford (1976). A calibration curve was established using bovine serum albumin as a standard at a concentration of 0.5 mg/mL.

Western immunoblotting analysis of recombinant *C. dromedarius* CYP1A

Recombinant fusion *C. dromedarius* CYP1A protein expressed in *E. coli* BL21(DE3) pLysS was detected by western blotting using a rabbit polyclonal antibody CYP1A1 (H-70, SC-20772, Santa Cruz Biotechnology, INC., USA) raised against amino acids 246-315 mapped onto an internal region of CYP1A1 of human origin. Total proteins were determined, and 50 μg was used for the immunoblotting analysis. Proteins were separated by electrophoresis on 12% SDS-polyacrylamide gel, and transferred to a nitrocellulose membrane (Laemmli, 1970; Towbin et al., 1979). After transfer, the membrane was blocked for 1 h with 50 mL 1X blocking buffer (Sigma-Aldrich, USA, Cat. No. B6429). Then, the membrane was first incubated with the primary anti-CYP1A1 antibody (1:500 dilution) for 8 h, and then washed for 30 min with 1X Tris buffer saline (1X TBS) containing 0.05% Tween-20. After a final wash with 1X TBS buffer, the membrane was incubated for 1 h with anti-rabbit secondary alkaline phosphatase labeled IgG antibody (1:2000 dilution). The membrane was washed again, and was finally developed using (nitro-blue tetrazolium salt/bromo-chloroindolyl-phosphate substrates. The protein bands were photographed by the Alpha Imager System (Alpha Innotech.Version: 2.0.0.9).


Structure modeling

Sequencing data were analyzed using BLASTn, and multiple sequence alignment was carried out using the ClustalW, BioEdit, and DNAStar programs. The CYP protein sequence was obtained by translating the sequenced DNA fragment using the Translation tool at the ExPasy server. The protein sequence was then submitted to the Swiss model server for structure prediction, and the structural data were analyzed by the PDB viewer program. Protein 3D structure models were built based on multiple-threading alignments by LOMET and iterative TASSER assembly simulation (Ortiz et al., 2002; Roy et al., 2010).

RESULTS

Cytochrome P450 1A, 2C, and 3A cDNA isolation and sequence analysis

A previous study reported that *C. dromedarius* expresses multiple P450 isoforms in hepatic and extrahepatic tissues and that most of these enzymes are constitutively expressed. In the previous study, the maximum expression levels of P450 proteins in the camel liver were in the order of P450 2E1 > 1A1 > 3A > 2B1/2. Moreover, detectable levels of the cytochrome P450 side-chain cholesterol cleavage and P450 21-hydroxylase were only observed in the camel testis (Alanazi et al., 2010). In the current study, full length camel P450 1A, P450 2C ,and P450 3A were obtained by RT-PCR using gene specific primers designed from the available EST camel genome project data base at http:// camel.kacst.edu.sa/. Utilizing these primers, specific PCR products were obtained that corresponded to 1563, 1473, and 1566 bp for P450 1A, 2C and 3A, respectively (Figure 1). The PCR products were cloned into pGEM-T-Easy vectorand the cDNA inserts were then sequenced. Nucleotide sequences corresponding to 1563, 1473, and 1566 bp were then compared with the nucleotide sequences deposited in the GenBank database using the Blastn program on the NCBI Blast server. The isolated cytochrome P450s showed high statistically significant similarity scores to numerous CYP450s from other species. The full-length cDNAs of *C. dromedarius* CYPs were deposited in the GenBank data base under the accession numbers of JQ619653 for CYP1A, JQ619655 for CYP3A, and JQ619654 for CYP2C. Alignment of the deduced amino acid sequences of *C. dromedarius* CYP 1A, 3A, and 2C with other species is shown in Figures 2-4. In the comparison of *C. dromedarius* CYP1A with other ungulates and human, the percentage identity was 83% for *Sus scrofa*, 82% for *Bos taurus*, 79% for *Macaca mulatta*, 78% for *Homo sapiens*, and 75% for *Mus musculus* (Figures 2-4 and Table 1). Meanwhile, *C. dromedarius* CYP3A showed 75% identity with *S. scrofa*, 72% with *H. sapiens*, and 68% with *M. musculus*. Moreover, *C. dromedarius* CYP2C showed 80% percentage identity with *B. taurus*, 74% with *M. mulatta*, 70% with *S. scrofa* and 68% with *H. sapiens*. The open reading frames of *C. dromedarius* CYP1A, 3A and 2C contained coding regions of 1563, 1566 and 1473 nucleotides, and the deduced amino acid sequences represented 520, 521 and 490 amino acids residues with predicted calculated molecular weights of 58.651, 58.594 and 56.030 kDa, respectively (Figures 5-7). Based on the amino acids composition, the predicted calculated isoelectric points using a computer algorithm were found to be 7.315 for CYP1A, 9.46 for CYP3A and 6.579 for CYP2C.
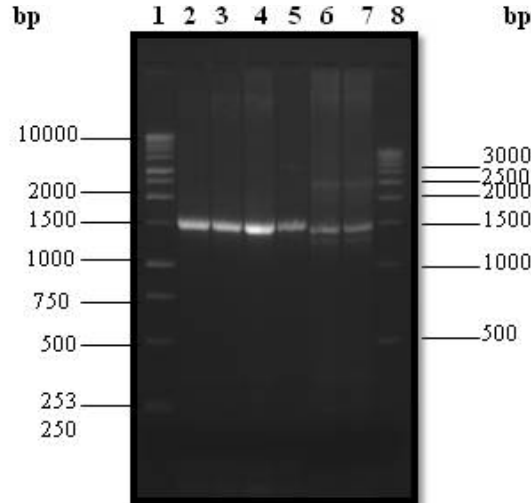
Figure 1. Agarose gel (1.0%) electrophoresis of PCR products for *Camelus dromedarius* CYP1A (*lanes 2* and *3*), CYP3A (*lanes 4* and *5*) and CYP 2C (*lanes 6* and *7*) cDNA. *lanes 1* and *8* represent 1- and 0.5-kbp DNA molecular weight markers.
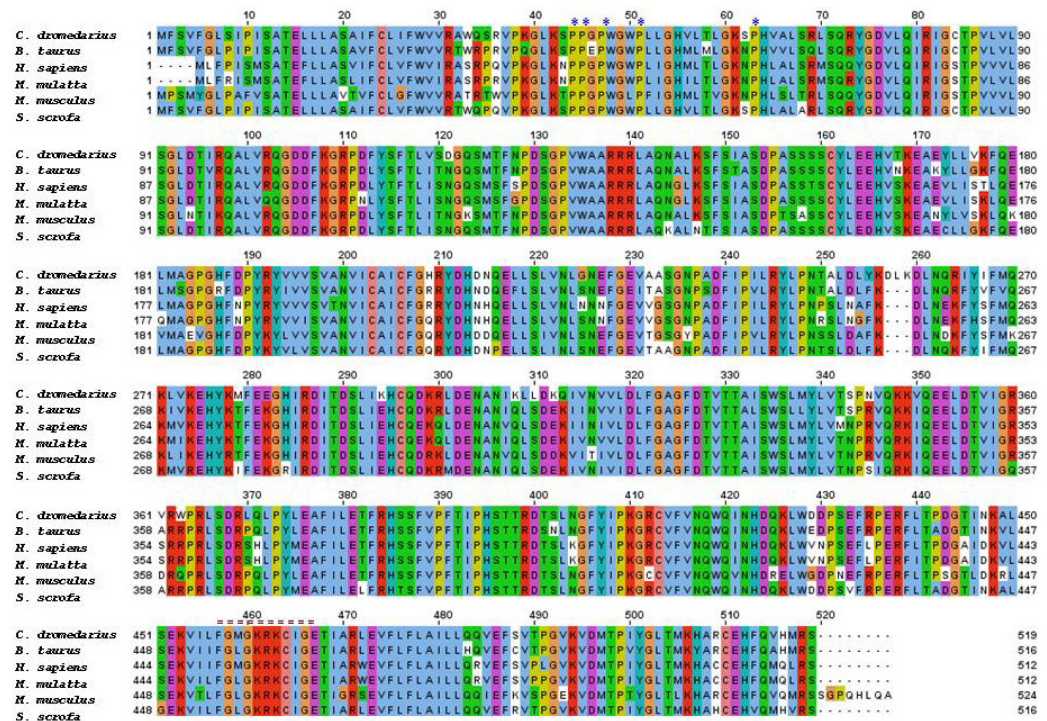
Figure 2. Alignment of the deduced amino acid sequences of *Camelus dromedarius* CYP1A (accession No. JQ619653), *Bos taurus* (accession No. DAA17555), *Homo sapiens* (accession No. P04798), *Macaca mulatta* (accession No. Q6GUR1), *Mus musculus* (accession No. P00184), and *Sus scrofa* (accession No. BAB85660).

Figure 3. Alignment of the deduced amino acid sequences of *Camelus dromedarius* CYP3A (accession No. JQ619654), *Homo sapiens* (accession No. AAA35744), *Mus musculus* (accession No. Q64464), and *Sus scrofa* (accession No. AAD04628).

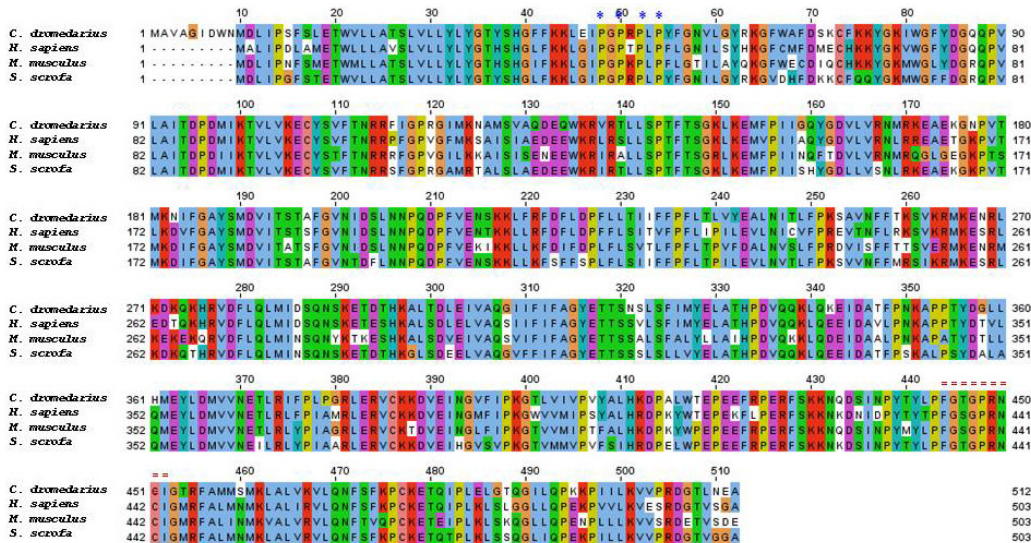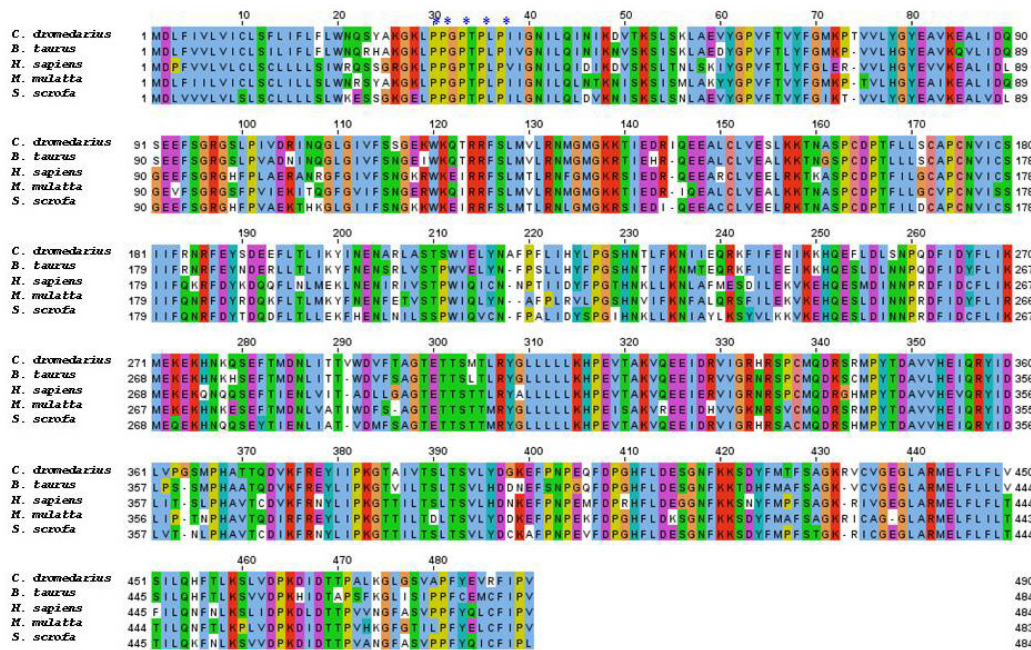**Figure 4.** Alignment of the deduced amino acid sequences of *Camelus dromedarius* CYP2C (accession No. JQ619655), *Bos taurus* (accession No. AAI53846), *Homo sapiens* (accession No. AAB59426), *Macaca mulatta* (accession No. ABB87189) and *Sus scrofa* (accession No. XP_001924722). Asterisks show the proline-rich regions and red-dashed lines show the conserved heme-binding sites.

Table 1. Homology of *Camelus dromedarius* CYPs with other species.

Animal species	CYP1A		CYP3A		CYP2C	
	Accession No.	Identity (%)	Accession No.	Identity (%)	Accession No.	Identity (%)
Sus scrofa	BAB85660	83	AAD04628	75	NP_001161307	70
Ovis aries	P56591	83	AAB02657	74	ADZ11091	79
Bos taurus	DAA17555	82	DAA15042	74	AAI53846	80
Equus caballus	XP_001493959	82	ABS52842	75	XP_001502043	78
Felis catus	Q5KQT7	81	BAK52240	73	XP_003994279	74
Otolemur garnettii	XP_003784563	81	XP_003798916	72	XP_003787365	71
Macaca mulatta	Q6GUR1	79	AAT49210	71	ABB87189	74
Homo sapiens	P04798	78	AAA35744	72	AAB59426	68
Mus musculus	P00184	78	Q64464	72	BAA04555	65

```
1     ATG TTC TCT GTG TTT GGA CTC TCC ATC CCC ATC TCG GCC ACG GAG   45
      M   F   S   V   F   G   L   S   I   P   I   S   A   T   E
46    CTT CTC CTG GCC TCT GCC ATC TTC TGC CTC ATA TTC TGG GTG GTC   90
      L   L   L   A   S   A   I   F   C   L   I   F   W   V   V
91    AGG GCC TGG CAG TCT CGG GTC CCC AAA GGC CTG AAG AGT CCA CCA   135
      R   A   W   Q   S   R   V   P   K   G   L   K   S   P   P
136   GGG CCC TGG GGC TGG CCC CTG CTC GGG CAC GTG CTG ACC TTG GGG   180
      G   P   W   G   W   P   L   L   G   H   V   L   T   L   G
181   AAG AGC CCA CAT GTG GCC CTG TCG CGG CTG AGC CAG CGC TAT GGA   225
      K   S   P   H   V   A   L   S   R   L   S   Q   R   Y   G
226   GAC GTG CTG CAG ATC CGC ATT GGC TGC ACG CCT GTA CTG GTG CTC   270
      D   V   L   Q   I   R   I   G   C   T   P   V   L   V   L
271   AGC GGC CTG GAC ACC ATC CGG CAG GCC CTG GTG CGG CAG GGC GAC   315
      S   G   L   D   T   I   R   Q   A   L   V   R   Q   G   D
316   GAT TTC AAG GGC CGG CCC GAC TTC TAC AGC TTC ACC CTG GTC TCT   360
      D   F   K   G   R   P   D   F   Y   S   F   T   L   V   S
361   GAT GGG CAG AGC ATG ACT TTC AAC CCA GAT TCT GGA CCA GTG TGG   405
      D   G   Q   S   M   T   F   N   P   D   S   G   P   V   W
406   GCT GCC CGC CGG CGC CTG GCC CAG AAT GCT CTG AAG AGT TTC TCC   450
      A   A   R   R   R   L   A   Q   N   A   L   K   S   F   S
451   ATC GCC TCC GAC CCG GCT TCC TCA TCC TCC TGC TAC CTG GAG GAG   495
      I   A   S   D   P   A   S   S   S   C   Y   L   E   E
496   CAT GTG ACC AAG GAG GCT GAG TAC CTC CTC GTT AAG TTT CAG GAG   540
      H   V   T   K   E   A   E   Y   L   L   V   K   F   Q   E
541   CTG ATG GCA GGG CCT GGG CAC TTT GAC CCC TAC AGG TAT GTA GTG   585
      L   M   A   G   P   G   H   F   D   P   Y   R   Y   V   V
586   GTG TCA GTG GCC AAT GTC ATC TGT GCT ATA TGC TTT GGC CAT CGC   630
      V   S   V   A   N   V   I   C   A   I   C   F   G   H   R
631   TAT GAC CAT GAC AAC CAA GAG CTG CTT AGC CTA GTC AAC CTT GGT   675
      Y   D   H   D   N   Q   E   L   L   S   L   V   N   L   G
676   AAC GAG TTT GGG GAG GTG GCT GCC TCT GGG AAC CCA GCC GAC TTT   720
      N   E   F   G   E   V   A   A   S   G   N   P   A   D   F
721   ATC CCT ATC CTC CGT TAC CTG CCC AAC ACT GCC CTG GAT CTC TAC   765
      I   P   I   L   R   Y   L   P   N   T   A   L   D   L   Y
766   AAG GAC CTG AAG GAC CTG AAT CAG AGG ATT TAC ATC TTC ATG CAG   810
      K   D   L   K   D   L   N   Q   R   I   Y   I   F   M   Q
811   AAG CTG GTC AAG GAA CAC TAT AAA ATG TTT GAG GAG GGC CAC ATC   855
      K   L   V   K   E   H   Y   K   M   F   E   E   G   H   I
856   CGG GAC ATC ACA GAC AGC CTG ATC AAA CAC TGT CAG GAC AAG AGG   900
      R   D   I   T   D   S   L   I   K   H   C   Q   D   K   R
901   CTG GAT GAG AAT GCC AAT ATC AAG CTG CTG GAT CAG CAG ATA GTT   945
      L   D   E   N   A   N   I   K   L   L   D   Q   Q   I   V
946   AAT GTC GTC CTC GAC CTC TTT GGA GCT GGG TTT GAC ACT GTC ACA   990
      N   V   V   L   D   L   F   G   A   G   F   D   T   V   T
991   ACT GCC ATC TCC TGG AGT CTC ATG TAC CTG GTG ACA AGC CCC AAT   1035
      T   A   I   S   W   S   L   M   Y   L   V   T   S   P   N
1036  GTG CAG AAA AAA GTC CAG GAG GAG CTG GAC ACA GTG ATT GGC AAG   1080
      V   Q   K   K   V   Q   E   E   L   D   T   V   I   G   K
1081  GGT TGC GGT GGC CTC GTG TCT CCG ACA GAC TTT CAG CTG CCC TAC   1125
      G   C   G   G   L   V   S   P   T   D   F   Q   L   P   Y
1126  TTG GAG GCC TTC ATC CTG GAG ACC TTC CGA CAC TCC TCC TTC GTC   1170
      L   E   A   F   I   L   E   T   F   R   H   S   S   F   V
1171  CCC TTC ACC ATC CCC CAC AGC ACC ACA AGA GAT ACA AGC CTG AAT   1215
      P   F   T   I   P   H   S   T   T   R   D   T   S   L   N
1216  GGC TTT TAC ATC CCC AAG GGG CGT TGT GTC TTT GTG AAC CAG TGG   1260
      G   F   Y   I   P   K   G   R   C   V   F   V   N   Q   W
1261  CAG ATC AAT CAT GAC CAG AAG CTG TGG GAT GAC CCA TCT GAG TTC   1305
      Q   I   N   H   D   Q   K   L   W   D   D   P   S   E   F
1306  CGA CCT GAA CGG TTT CTT ACC CCT GAT GGC ACC ATC AAC AAG GCA   1350
      R   P   E   R   F   L   T   P   D   G   T   I   N   K   A
1351  CTG AGT GAG AAG GTG ATT CTC TTC GGC ATG GGC AAG CGG AAG TGC   1395
      L   S   E   K   V   I   L   F   G   M   G   K   R   K   C
1396  ATT GGT GAG ACC ATT GCC CGC GTG GTC TTT CTC TTC CTG GCC   1440
      I   G   E   T   I   A   R   V   V   F   L   F   L   A
1441  ATC CTG CTG CAG CAG GTG GAA TTC AGT GTG ACC CCA GGT GTG AAG   1485
      I   L   L   Q   Q   V   E   F   S   V   T   P   G   V   K
1486  GTG GAC ATG ACC CCC ATC TAT GGG CTG ACC ATG AAG CAT GCC CGC   1530
      V   D   M   T   P   I   Y   G   L   T   M   K   H   A   R
1531  TGT GAG CAC TTC CAG GTG CAT ATG CGC TCT TAG   1563
      C   E   H   F   Q   V   H   M   R   S   *
```

Figure 5. Nucleotides and deduced amino acid sequences of *Camelus dromedarius* CYP1A. Proline-rich regions and heme-binding sites are underlined in red.

```
1     ATG GCG GTC GCG GGA ATC GAT TGG AAC ATG GAC CTG ATC CCA AGC   45
      M   A   V   A   G   I   D   W   N   M   D   L   I   P   S
46    TTT TCC TTA GAA ACC TGG GTT CTC CTG GCT ACC AGC CTG GTG CTG   90
      F   S   L   E   T   W   V   L   L   A   T   S   L   V   L
91    CTC TAT CTA TAT GGG ACT TAT TCA CAT GGA TTT TTT AAG AAG CTG  135
      L   Y   L   Y   G   T   Y   S   H   G   F   F   K   K   L
136   GAG ATT CCT GGG CCA AGA CCT CTG CCT TAT TTT GGA AAC GTC CTG  180
      E   I   P   G   P   R   P   L   P   Y   F   G   N   V   L
181   GGC TAC CGA AAG GGT TTT TGG GCT TTT GAC AGT AAA TGT TTT AAA  225
      G   Y   R   K   G   F   W   A   F   D   S   K   C   F   K
226   AAG TAT GGA AAA ATA TGG GGG TTT TAT GAC GGT CAA CAG CCT GTG  270
      K   Y   G   K   I   W   G   F   Y   D   G   Q   Q   P   V
271   TTG GCT ATC ACA GAT CCA GAC ATG ATT AAA ACA GTA CTA GTG AAA  315
      L   A   I   T   D   P   D   M   I   K   T   V   L   V   K
316   GAA TGT TAT TCT GTC TTC ACA AAT CGG AGG TTC ATT GGC CCA AGG  360
      E   C   Y   S   V   F   T   N   R   R   F   I   G   P   R
361   GGA ATT ATG AAA AAT GCC ATG TCT GTG GCT CAG GAC GAA CAA TGG  405
      G   I   M   K   N   A   M   S   V   A   Q   D   E   Q   W
406   AAG AGA GTA CGA ACA CTG CTG TCT CCA ACC TTC ACC AGT GGA AAG  450
      K   R   V   R   T   L   L   S   P   T   F   T   S   G   K
451   CTC AAG GAG ATG TTC CCC ATC ATT GGT CAG TAT GGA GAT GTG TTG  495
      L   K   E   M   F   P   I   I   G   Q   Y   G   D   V   L
496   GTG AGA AAT ATG AGG AAG GAA GCA GAG AAA GGC AAC CCC GTC ACC  540
      V   R   N   M   R   K   E   A   E   K   G   N   P   V   T
541   ATG AAG AAC ATC TTT GGG GCC TAC AGC ATG GAT GTG ATT ACC AGC  585
      M   K   N   I   F   G   A   Y   S   M   D   V   I   T   S
586   ACA GCA TTT GGA GTG AAC ATC GAT TCC CTC AAC AAC CCA CAA GAT  630
      T   A   F   G   V   N   I   D   S   L   N   N   P   Q   D
631   CCC TTT GTG GAA AAT TCC AAG AAA CTC TTT AGA TTT GAT TTC CTT  675
      P   F   V   E   N   S   K   K   L   F   R   F   D   F   L
676   GAT CCA TTC CTT CTC ACA ATA ATA TTC TTT CCA TTC CTC ACC CTA  720
      D   P   F   L   L   T   I   I   F   F   P   F   L   T   L
721   GTC TAC GAA GCA TTA AAT ATT ACT CTG TTT CCA AAA AGT GCT GTG  765
      V   Y   E   A   L   N   I   T   L   F   P   K   S   A   V
766   AAT TTT TTC ACA AAG TCC GTA AAA AGA ATG AAG GAA AAT CGA CTC  810
      N   F   F   T   K   S   V   K   R   M   K   E   N   R   L
811   AAA GAT AAA CAA AAG CAC CGA GTG GAT TTT CTT CAG CTG ATG ATT  855
      K   D   K   Q   K   H   R   V   D   F   L   Q   L   M   I
856   GAC TCC CAG AAT TCC AAA GAA ACA GAC ACC CAT AAA GCT CTG ACT  900
      D   S   Q   N   S   K   E   T   D   T   H   K   A   L   T
901   GAC CTA GAA ATA GTG GCC CAA GGC ATT ATC TTT ATT TTT GCT GGC  945
      D   L   E   I   V   A   Q   G   I   I   F   I   F   A   G
946   TAT GAG ACC ACC AGC AAT TCT CTC TCC TTC ATT ATG TAT GAA CTG  990
      Y   E   T   T   S   N   S   L   S   F   I   M   Y   E   L
991   GCC ACG CAC CCC GAT GTC CAG CAG AAG CTA CAG AAG GAG ATC GAT 1035
      A   T   H   P   D   V   Q   Q   K   L   Q   K   E   I   D
1036  GCG ACT TTC CCC AAC AAG GCG CCC ACC TAC GAT GGC CTG CTA     1080
      A   T   F   P   N   K   A   P   P   T   Y   D   G   L   L
1081  CAC ATG GAG TAT CTT GAC ATG GTG GTG AAT GAA ACT CTC AGA ATA 1125
      H   M   E   Y   L   D   M   V   V   N   E   T   L   R   I
1126  TTC CCA CTT CCT GGG AGA CTG GAG AGG GTC TGT AAG AAA GAT GTG 1170
      F   P   L   P   G   R   L   E   R   V   C   K   K   D   V
1171  GAA ATC AAT GGG GTG TTC ATT CCC AAA GGG ACA CTG GTG ATC GTG 1215
      E   I   N   G   V   F   I   P   K   G   T   L   V   I   V
1216  CCA GTC TAC GCT CTT CAC AAA GAC CCA GCA CTT TGG ACA GAG CCT 1260
      P   V   Y   A   L   H   K   D   P   A   L   W   T   E   P
1261  GAG GAG TTC CGT CCG GAA AGG TTC AGC AAG AAC AAC CAG GAC AGC 1305
      E   E   F   R   P   E   R   F   S   K   K   N   Q   D   S
1306  ATA AAT CCT TAC ACA TAC CTG CCT TTT GGA ACT GGC CCC CGA AAC 1350
      I   N   P   Y   T   Y   L   P   F   G   T   G   P   R   N
1351  TGC ATC GGC ACG AGG TTT GCT ATG ATG AGC ATG AAA CTT GCT CTC 1395
      C   I   G   T   R   F   A   M   M   S   M   K   L   A   L
1396  GTC AAA GTC CTG CAG AAC TTC TCC TTC AAA CCT TGC AAA GAA ACA 1440
      V   K   V   L   Q   N   F   S   F   K   P   C   K   E   T
1441  CAG ATC CCC CTG GAA TTA GGC ACT CAA GGC ATT TTG CAA CCG AAA 1485
      Q   I   P   L   E   L   G   T   Q   G   I   L   Q   P   K
1486  AAA CCC ATC ATT TTA AAA GTT GTG CCC CGA GAC GGA ACC TTA AAT 1530
      K   P   I   I   L   K   V   V   P   R   D   G   T   L   N
1531  GAA GCC TGA CTT TCC CTA AGG ACT TCG AAA ATC ACT     1566
      E   A   *   L   S   L   R   T   S   K   I   T
```

Figure 6. Nucleotides and deduced amino acid sequences of *Camelus dromedarius* CYP3A. Proline-rich regions and heme-binding sites are underlined in red.

```
1     ATG GAT CTC TTC ATA GTC CTG GTG ATT TGT CTT TCT TTT TTG ATT   45
      H   D   L   F   I   V   L   V   I   C   L   S   F   L   I
46    TTC CTT TTT CTG TGG AAC CAA AGC TAT GCC AAA GGG AAG CTG CCA   90
      F   L   F   L   W   N   Q   S   Y   A   K   G   K   L   P
91    CCT GGC CCT ACC CCT CTC CCC ATC ATT GGG AAT ATT CTA CAG ATA   135
      P   G   P   T   P   L   P   I   I   G   N   I   L   Q   I
136   AAC ATT AAG GAT GTC ACC AAA TCC CTG AGC AAG CTA GCA GAA GTC   180
      N   I   K   D   V   T   K   S   L   S   K   L   A   E   V
181   TAT GGC CCT GTG TTC ACT GTA TAT TTT GGC ATG AAG CCC ACC GTG   225
      Y   G   P   V   F   T   V   Y   F   G   M   K   P   T   V
226   GTG TTA TAC GGA TAT GAA GCA GTG AAG GAA GCT TTG ATT GAT CAG   270
      V   L   Y   G   Y   E   A   V   K   E   A   L   I   D   Q
271   AGT GAG GAG TTC TCT GGC CGA GGC AGT TTA CCA ATA GTA GAC AGA   315
      S   E   E   F   S   G   R   G   S   L   P   I   V   D   R
316   ATC AAC CAG GGA TTA GGA ATT GTT TTC AGC AGT GGA GAA AAA TGG   360
      I   N   Q   G   L   G   I   V   F   S   S   G   E   K   W
361   AAG CAA ACC CGG CGT TTC TCC CTC ATG GTT TTG CGG AAT ATG GGG   405
      K   Q   T   R   R   F   S   L   H   V   L   R   N   H   G
406   ATG GGG AAG AAG ACT ATT GAA GAT AGA ATT CAA GAG GAA GCC TTG   450
      H   G   K   K   T   I   E   D   R   I   Q   E   E   A   L
451   TGT TTG GTA GAA TCA TTA AAA AAA ACT AAT GCA TCT CCG TGC GAT   495
      C   L   V   E   S   L   K   K   T   N   A   S   P   C   D
496   CCA ACT TTC CTT CTG AGC TGT GCT CCC TGC AAT GTG ATC TGC TCT   540
      P   T   F   L   L   S   C   A   P   C   N   V   I   C   S
541   ATA ATC TTC CGG AAC CGT TTT GAG TAC AGT GAT GAG GAA TTC CTA   585
      I   I   F   R   N   R   F   E   Y   S   D   E   E   F   L
586   ACT TTG ATA AAA TAT ATC AAT GAA AAT GCA AGA CTT GCA AGC ACC   630
      T   L   I   K   Y   I   N   E   N   A   R   L   A   S   T
631   TCC TGG ATA GAG CTC TAC AAT GCT TTC CCC TTT TTA ATA CAT TAT   675
      S   W   I   E   L   Y   N   A   F   P   F   L   I   H   Y
676   CTC CCA GGA AGT CAT AAT ACA CTA TTT AAA AAC ATA ATT GAG CAA   720
      L   P   G   S   H   N   T   L   F   K   N   I   I   E   Q
721   AGA AAG TTC ATT TTT GAG AAC ATA AAG AAA CAC CAA GAA TTC CTT   765
      R   K   F   I   F   E   N   I   K   K   H   Q   E   F   L
766   GAC CTC AGT AAC CCT CAA GAC TTT ATT GAT TAC TTC CTG ATT AAA   810
      D   L   S   N   P   Q   D   F   I   D   Y   F   L   I   K
811   ATG GAA AAG GAA AAG CAC AAT AAG CAG TCT GAA TTT ACC ATG GAC   855
      H   E   K   E   K   H   N   K   Q   S   E   F   T   H   D
856   AAC TTG ATC ACT ACT GTA TGG GAT GTA TTT ACT GCT GGA ACA GAG   900
      N   L   I   T   T   V   W   D   V   F   T   A   G   T   E
901   ACA ACA AGC ATG ACT TTG AGA TAT GGA CTC CTG CTC CTG CTG AAG   945
      T   T   S   H   T   L   R   Y   G   L   L   L   L   L   K
946   CAC CCA GAG GTC ACA GCT AAG GTC CAG GAA GAG ATT GAC CGT GTG   990
      H   P   E   V   T   A   K   V   Q   E   E   I   D   R   V
991   ATT GGC AGA CAC CGG AGC CCC TGC ATG CAG GAC AGG AGC CGC ATG   1035
      I   G   R   H   R   S   P   C   M   Q   D   R   S   R   H
1036  CCC TAC ACG GAT GCC GTG GTG CAC GAA ATC CAG AGA TAC ATC GAC   1080
      P   Y   T   D   A   V   V   H   E   I   Q   R   Y   I   D
1081  CTG GTG CCC GGC AGC ATG CCC CAT GCA ACA ACT CAG GAC GTT AAA   1125
      L   V   P   G   S   H   P   H   A   T   T   Q   D   V   K
1126  TTC AGA GAA TAC ATT ATT CCC AAG GGC ACA GCC ATA GTA ACA TCT   1170
      F   R   E   Y   I   I   P   K   G   T   A   I   V   T   S
1171  CTG ACT TCT GTT CTG TAT GAT GGC AAA GAG TTT CCC AAT CCA GAG   1215
      L   T   S   V   L   Y   D   G   K   E   F   P   N   P   E
1216  CAG TTT GAC CCT GGT CAC TTC CTG GAT GAA AGT GGC AAC TTT AAG   1260
      Q   F   D   P   G   H   F   L   D   E   S   G   N   F   K
1261  AAG AGT GAC TAC TTC ATG ACT TTT TCA GCA GGA AAA AGA GTT TGT   1305
      K   S   D   Y   F   H   T   F   S   A   G   K   R   V   C
1306  GTT GGA GAA GGC CTG GCC CGC ATG GAG CTG TTT TTA TTC CTG GTC   1350
      V   G   E   G   L   A   R   H   E   L   F   L   F   L   V
1351  AGC ATT TTA CAG CAT TTC ACT TTG AAA TCT CTG GTT GAT CCA AAA   1395
      S   I   L   Q   H   F   T   L   K   S   L   V   D   P   K
1396  GAC ATT GAC ACT ACC CCA GCT CTC AAA GGG TTA GGC TCT GTA GCA   1440
      D   I   D   T   T   P   A   L   K   G   L   G   S   V   A
1441  CCC TTC TAT GAG GTC CGT TTC ATT CCA GTC TGA
      P   F   Y   E   V   R   F   I   P   V   *
```

Figure 7. Nucleotides and deduced amino acid sequences of *Camelus dromedarius* CYP2C. Proline-rich regions and heme-binding sites are underlined in red.

Predicted 3D structure of *C. dromedarius* CYP1A, 3A and 2C

The Arabian camel CYP1A, 3A and 2C proteins motifs secondary structures annotation prediction showed several common characteristic features for CYPs (Figures 8-10). First, the N-terminal hydrophobic segment or signal anchor segment (amino acid residues 1-38). Second, following the N-terminal signal peptide there is a proline-rich region (4-5 proline residues). Third, there is a highly conserved heme-binding cysteine-containing peptide that is localized near the C-terminal end of the CYP proteins that forms a pocket in the 3D structure of the CYPs. This cysteine-containing pocket contains up to 10 amino acid residues (Figure 11), including the invariant cysteine residue that is involved in the ligation of the heme iron prosthetic group. Figure 5 shows the alignment of *C. dromedarius* CYP1A, 3A and 2C sequences with the CYP sequences of other species around the heme-binding site. Comparing the 3 camels CYP heme-binding signatures revealed the conservation of the invariant amino acid residues (phenylalanine, glycine, and cysteine) in the 3 isolated CYPs cDNA (Figure 11). Proteins with similar amino acid sequences have a tendency to adopt similar 3D structures. Camel 3D CYP1A, 3A, and 2C protein crystal structural models were predicted from amino acid sequences using the I-TASSER server based on multiple-threading alignments by LOMETS and iterative TASSER assembly simulations to obtain the most precise 3D structural prediction. Figure 12 show the overall fold and secondary structure contents of the camel CYP 1A, 3A and 2C as ribbon diagrams. The 3 camel CYPs structures clearly contain the same set of secondary structures elements in very similar folds, with respect to the positions of alpha helices, β-strands, and the highly conserved β-bulge or cysteine-pocket. The heme moiety of the CYPs is covalently bound to the invariant cysteine (Cys[465] for CYP1A, Cys[451] for CYP3A, and Cys[435] for CYP2C), which is found in the β-bulge region called the cysteine pocket.
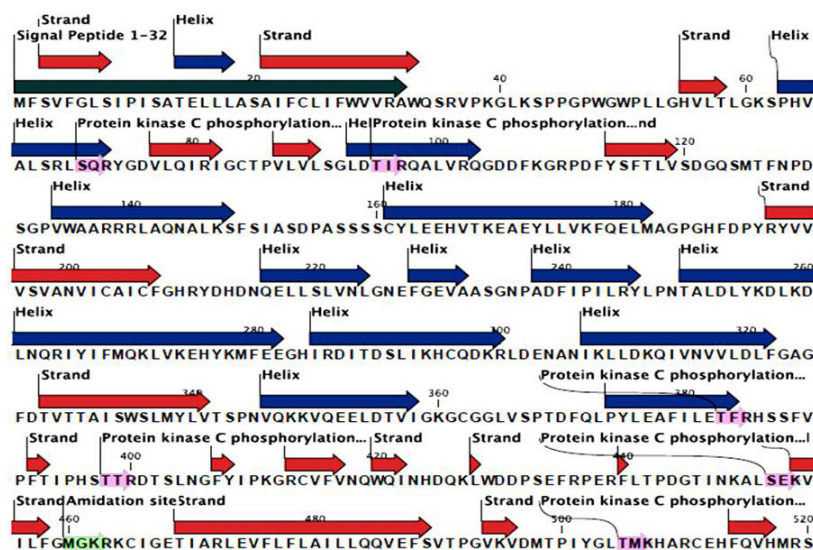
Figure 8. Arabian camel CYP1A protein motifs secondary structure annotation prediction showing the N-terminal signal peptide, proline-rich region, heme-binding site and protein kinase C-phosphorylation sites, along with secondary structure prediction.
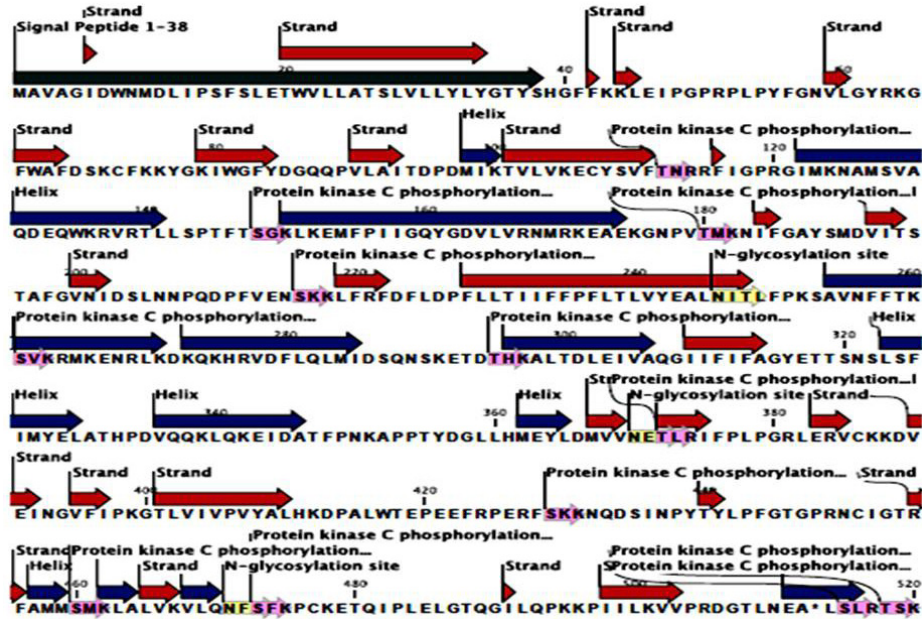
Figure 9. Arabian camel CYP3A protein motifs secondary structure annotation prediction showing the N-terminal signal peptide, proline-rich region, heme-binding site and protein kinase C-phosphorylation sites along with secondary structure prediction.

Figure 10. Arabian camel CYP2C protein motifs secondary structure annotation prediction showing the N-terminal signal peptide, proline-rich region, heme-binding site and protein kinase C-phosphorylation sites along with secondary structure prediction.
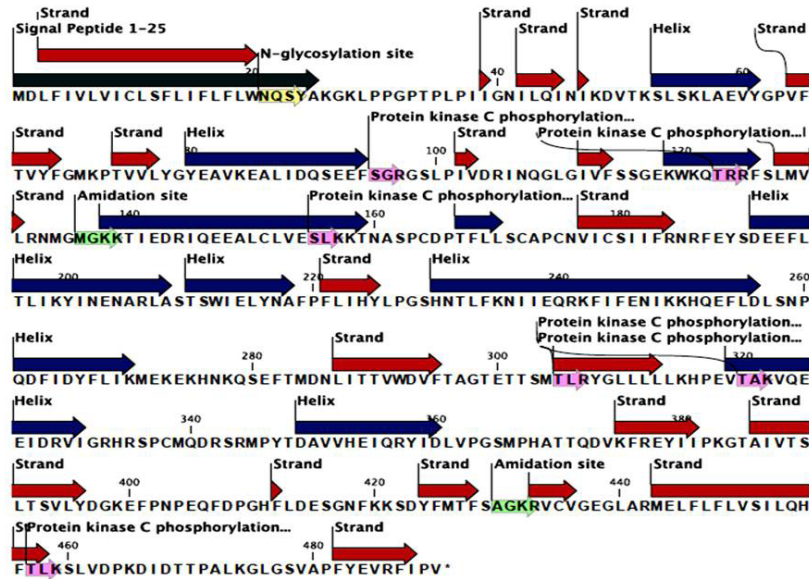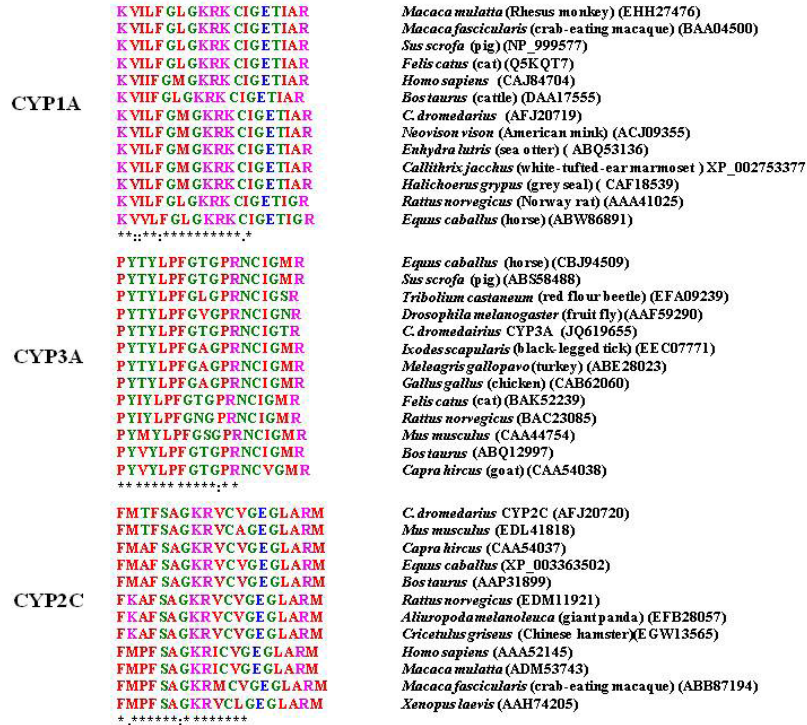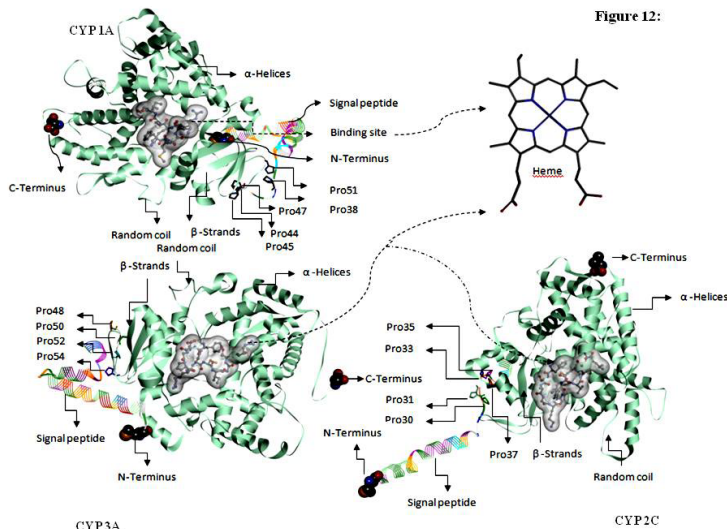
CYP1A

Sequence	Species
KVILF GLGKRK CIGETIAR	*Macaca mulatta* (Rhesus monkey) (EHH27476)
KVILF GLGKRK CIGETIAR	*Macaca fascicularis* (crab-eating macaque) (BAA04500)
KVILF GLGKRK CIGETIAR	*Sus scrofa* (pig) (NP_999577)
KVILF GLGKRK CIGETIAR	*Felis catus* (cat) (Q5KQT7)
KVIIF GMGKRK CIGETIAR	*Homo sapiens* (CAJ84704)
KVIIF GLGKRK CIGETIAR	*Bos taurus* (cattle) (DAA17555)
KVILF GMGKRK CIGETIAR	*C. dromedarius* (AFJ20719)
KVILF GMGKRK CIGETIAR	*Neovison vison* (American mink) (ACJ09355)
KVILF GMGKRK CIGETIAR	*Enhydra lutris* (sea otter) (ABQ53136)
KVILF GMGKRK CIGETIAR	*Callithrix jacchus* (white-tufted-ear marmoset) XP_002753377
KVILF GMGKRK CIGETIAR	*Halichoerus grypus* (grey seal) (CAF18539)
KVILF GLGKRK CIGETIGR	*Rattus norvegicus* (Norway rat) (AAA41025)
KVVLF GLGKRK CIGETIGR	*Equus caballus* (horse) (ABW86891)
::**:*********** *	

CYP3A

Sequence	Species
PYTYLPFGTGPRNCIGMR	*Equus caballus* (horse) (CBJ94509)
PYTYLPFGTGPRNCIGMR	*Sus scrofa* (pig) (ABS58488)
PYTYLPFGLGPRNCIGSR	*Tribolium castaneum* (red flour beetle) (EFA09239)
PYTYLPFGVGPRNCIGNR	*Drosophila melanogaster* (fruit fly) (AAF59290)
PYTYLPFGTGPRNCIGTR	*C. dromedairius* CYP3A (JQ619655)
PYTYLPFGAGPRNCIGMR	*Ixodes scapularis* (black-legged tick) (EEC07771)
PYTYLPFGAGPRNCIGMR	*Meleagris gallopavo* (turkey) (ABE28023)
PYTYLPFGAGPRNCIGMR	*Gallus gallus* (chicken) (CAB62060)
PYIYLPFGTGPRNCIGMR	*Felis catus* (cat) (BAK52239)
PYIYLPFGNGPRNCIGMR	*Rattus norvegicus* (BAC23085)
PYMYLPFGSGPRNCIGMR	*Mus musculus* (CAA44754)
PYVYLPFGTGPRNCIGMR	*Bos taurus* (ABQ12997)
PYVYLPFGTGPRNCVGMR	*Capra hircus* (goat) (CAA54038)
** ***** *****:* *	

CYP2C

| Sequence	Species
FMTFSAGKRVCVGEGLARM	*C. dromedarius* CYP2C (AFJ20720)
FMTFSAGKRVCAGEGLARM	*Mus musculus* (EDL41818)
FMAFSAGKRVCVGEGLARM	*Capra hircus* (CAA54037)
FMAFSAGKRVCVGEGLARM	*Equus caballus* (XP_003363502)
FMAFSAGKRVCVGEGLARM	*Bos taurus* (AAP31899)
FKAFSAGKRVCVGEGLARM	*Rattus norvegicus* (EDM11921)
FKAFSAGKRVCVGEGLARM	*Ailuropoda melanoleuca* (giant panda) (EFB28057)
FKAFSAGKRVCVGEGLARM	*Cricetulus griseus* (Chinese hamster)(EGW13565)
FMPFSAGKRICVGEGLARM	*Homo sapiens* (AAA52145)
FMPFSAGKRICVGEGLARM	*Macaca mulatta* (ADM53743)
FMPFSAGKRMCVGEGLARM	*Macaca fascicularis* (crab-eating macaque) (ABB87194)
FMPFSAGKRVCLGEGLARM	*Xenopus laevis* (AAH74205)
* :******:* *******	

Figure 11. Alignment of the *Camelus dromedarius* CYP1A, CYP3A, and CYP2C, with the sequences of other species around the heme-binding site. Asterisks indicate conserved invariant residues; dots and semicolons denote variant residues.


Figure 12. Predictive 3D structure models of the *Camelus dromedarius* CYP1A, CYP3A, and CYP2C, showing the N-terminal signal peptide, proline-rich region and the heme-binding pocket.

Cloning and expression of *C. dromedarius* CYP1A cDNA in *E. coli*

The full length open reading frame of *C. dromedarius* CYP1A cDNA was cloned into the pET28a(+) vector utilizing the *Bam*HI and *Not*I restriction sites, and expressed in *E. coli* (BL21)DE3pLysS cells. A high level of recombinant fusion CYP1A was attained by adding 1.0 mM IPTG which induced the T7 promoter after an 18-h incubation period (Figure 13). The result of Western immunoblotting analysis (Figure 14) indicated that the anti-human CYP1A1 antibody cross reacted with the *C. dromedarius* recombinant CYP1A fusion protein to produce a specific cross reactivity band of about 60.0 kDa, which corresponds to the His tag fusion protein.
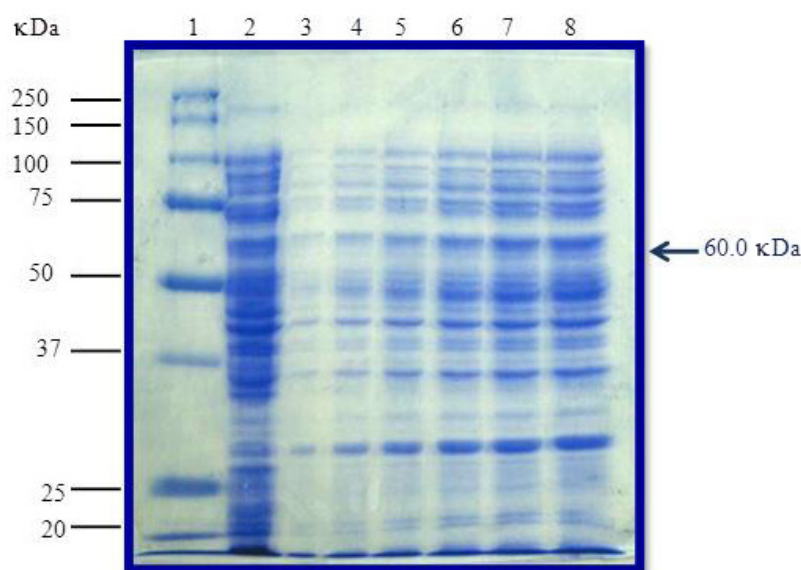
Figure 13. Sodium dodecyl sulfate gel (12%) electrophoresis for recombinant *Camelus dromedarius* cytochrome P450 1A in *Escherichia coli* BL21(DE3)pLysS (pET28a+) induced with 1 mM IPTG (*lanes 3-8*) and uninduced culture (*lane 2*). *Lane 1* represents Precision Plus pre-stained protein standards (BioRad Cat No. 161-0373).
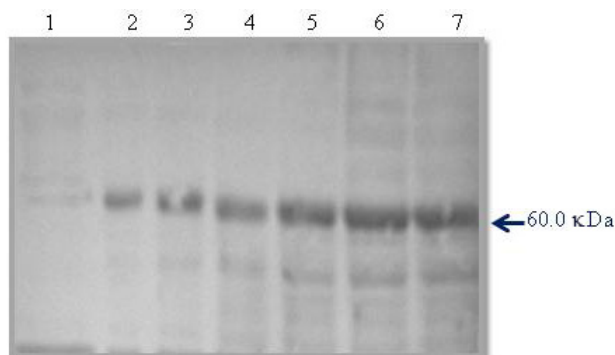
Figure 14. Immunoblotting of recombinant *Camelus dromedarius* CYP1A1 with anti-cytochrome P450 1A1 antibody. *Lanes 2-7* = different concentrations of recombinant protein from 10 to 60 µg of the induced culture proteins. *Lane 1* = 60 µg uninduced culture.

DISCUSSION

Despite the economic, cultural, and biological importance of the Arabian camel, information remains limited about this species' Phase I drug metabolizing enzymes. The molecular characterization of Arabian camel CYP450 enzymes is very important for understanding the impact of exposure to various environmental factors on the health status of this unique animal. RT-PCR and immunoblotting results from a previous study (Alanazi et al., 2010) of *C. dromedarius* CYPs demonstrated that significant mRNA and protein of CYPs isoforms are found in camel liver and extrahepatic tissues including the lung, kidney, spleen, and testis. The present study focused on the molecular characterization of 3 important cytochrome P450s mainly, CYP1A, 3A and 2C from the Arabian camel. Cytochrome P450 1A1 is a member of a multigene family of xenobiotic metabolizing enzymes, and plays a physiological role in the detoxification of polycyclic aromatic compounds (PAHs). However, the activity of this enzyme may be deleterious, because it generates mutagenic metabolites and active oxygen (Mimura and Fujii-Kuriyama, 2003). In this gene, the induction of the CYP1A1 isoenzyme by certain environmental chemicals such as 2, 3, 7, 8-tetrachlorodibenzo-ρ-dioxin and polymorphism primarily account for its susceptibility to PAH-induced carcinogenesis, including lung cancer (Shah et al., 2009). CYP3A isoenzymes are the major CYPs expressed in humans, and are involved in the metabolism of majority of the drugs. Besides metabolizing drugs, these enzymes also play an important role in the metabolism of steroids and toxins, including carcinogens (Kirby et al., 1993; Gonzalez and Gelboin, 1994; Evans and Relling, 1999). Although the expression of CYP3A isoenzymes is highest in the liver, their expression has also been demonstrated in extrahepatic tissues (Michael and Doherty, 2005). Variability in the expression of CYP3A enzyme might also affect an individual's susceptibility to cancers caused by environmental procarcinogens, which are CYP3A substrates (Kirby et al., 1993). The cytochrome P450 2C subfamily is predominantly expressed in the liver, and contains 4 highly homologous genes 2C8, 2C9, 2C18, and 2C19, which are responsible for the metabolism of approximately 20% of all clinically used drugs (Michael and Doherty, 2005).

In the current study, we showed the amplification of *C. dromedarius* CYP1A, 2C and 3A full-length cDNA using primer sets spanning the entire open reading frame, with cDNA fragments with the anticipated sizes of 1563, 1473, and 1566 bp (Figures 1-4) being exclusively obtained. The cDNA sequences contained 520, 490, and 521 amino acid residues in the proteins of 58.651, 56.03 and 58.594 kDa, with these CYP450 enzymes being matched with several CYPs sequences in the GenBank data base (Figures 2-4). Several observations from the primary structures and from the multiple sequence alignments (Figures 2-10) of camel CYPs are of interest.

First, the primary sequence homology between camel CYPs and other species was greater than 70% (Table 1). Second, the primary structure (Figures 5-10) contained the characteristic N-terminal hydrophobic segment or signal peptide that is rich in hydrophobic amino acid residues, such as valine, leucine, isoleucine, and alanine. This signal peptide is a characteristic feature of P450 cytochromes which are membrane bound proteins with the exception of bacterial enzymes. Microsomal CYP enzymes are tethered to the membrane through this hydrophobic transmembrane helix at the N-terminus of the protein which also serves as a targeting sequence for the signal recognition particle dependent co-translational incorporation of a nascent CYP protein into the endoplasmic reticulum membrane (Bar-Nun et al., 1980; Sakaguchi et al., 1984). About 90% of hepatic cytochrome P450 enzymes are present in micro-

somal membranes, with the remainder being found in the mitochondria (Nelson et al., 2004).

Third, following the N-terminal hydrophobic segment, there is a proline-rich region (Pro[44, 45, 47, 51] in CYP1A; Pro[48, 50, 52, 54] in CYP3A, and Pro[30, 31, 33, 35, 37] in CYP2C) that acts as an alpha helix breaker in the folding of the CYP proteins. Fourth, there is a highly conserved heme-binding cysteine-containing peptide that is localized near the C-terminal end of the proteins (Figures 2-11). This cysteine-containing pocket is a common CYP450 signature (Guengerich, 1997), and this motif contains up to 10 amino acid residues, including the invariant cysteine residue (Cys[465] in CYP1A, Cys[451] in CYP3A, and Cys[435] in CYP2C) which is involved in the ligation of the heme-iron prosthetic group. In general, this signature takes the form of FXXGXXXCXG (Figure 11), where F is phenylalanine, G is glycine, C is cysteine, and X is any amino acid. Fifth, a dilysine motif KKXX, (where K is the lysine residue and X is any other amino acid) at the C-terminal end of the cytoplasmic domain was characterized in camel CYP2C (Lys[420, 421]) and 3A (Lys[495, 496]), but it was absent in CYP1A. The dilysine motif, KKXX, facilitates the exclusion of endoplasmic reticulum microsomal CYP enzymes from being transported through the Golgi to other cell compartments, or from exocytosis (Andersson et al., 1999).

The 3D-structure of a protein provides valuable insights about its functions. Ideally, experimental tools and techniques, such as X-ray crystallography nuclear magnetic resonance (NMR) spectroscopy and electron microscopy, are used to determine the 3D-structure of proteins. Unfortunately, these techniques cannot be used on the vast majority of proteins, including microsomal cytochrome P450s, as the proteins are difficult to crystallize, insufficiently soluble, or too large for NMR studies. Alternative methods have been developed to determine the 3D-structures of these proteins, such as comparative (or homology) modeling. It is possible to predict the 3D structure of a protein based solely on knowledge of its amino acid sequences. Hence, we generated a 3D structure of the putative *C. dromedarius* CYP1A, 3A, and 2C, for which the nucleotide and amino acid sequences (but not the 3D structures) are known. The overall secondary structures, folding, and topology are quite similar (Figure 12). Common orientation and anchoring in the membrane by an N-terminal helix was observed for all camel CYPs under study. The heme moiety of the camel CYPs is covalently bound to the invariant Cysteine (Cys[465] for CYP1A, Cys[451] for CYP3A, and Cys[435] for CYP2C) found in the β-bulge region called the Cys-pocket. This β-bulge appears to have the role of enveloping the invariant cysteine residue in a hydrophobic environment (Hasemann et al., 1995). Three residues besides the cysteine (Gly[461, 467] and Phe[458] in 1A, Gly[447, 453] and Phe[444] in 3A, and Gly[431, 437] and Phe[428] in 2C) are very strictly conserved among most of the CYP450 proteins. These 2 glycine residues are involved in the formation of the β-hairpin turn and facilitate the sharp turn from the Cys-pocket into the L-helix, and are involved in determining the proximity to the heme. The phenylalanine side chain completes the hydrophobic enclosure of the proximal heme in combination with other side-chains and the main chain atoms of the Cys-pocket (Kassner, 1973; Yasukochi et al., 1994). Substrates bind to a cavity above the heme surface, because the heme forms the base of the active site cavity, and it must be positioned close to the reactive iron-oxo intermediate for catalysis (Hasemann et al., 1995). The regions that form the outer surfaces of the substrate binding cavity are generally more divergent among enzymes compared to other parts of the proteins, leading to differences in the size, shape, and chemical features of the active sites that provide discrimination for different substrates.

A pilot recombinant protein expression study using *E. coli* BL21(DE3) pLysS harboring pET28a(+) carrying the full length CYP1A cDNA indicated that recombinant fusion camel

CYP1A protein cross-reacted with the anti-human cytochrome P450 1A antibody (Figure 14). The cross reactivity band was found to be at a molecular weight of 60 kDa, which represents the size of the CYP1A protein in addition to a 1 kDa His-tag fusion peptide localized at the N-terminus of the recombinant fusion protein. Immunoreactivity was not detected in the un-induced *E. coli* BL21(DE3) pLysS (Figures 13, 14).

CONCLUSIONS

The molecular characterization of the Arabian camel CYP1A, 3A, and 2C revealed a high degree of similarity with other ungulates and human. The data presented in this study provides baseline information on which to develop further studies of camel CYP proteins, such as the characterization of the purified native and recombinant proteins and the role of camel CYPs, especially the CYP3A in the metabolism of xenobiotic compounds and environmental pollutants. Therefore, it would be useful to improve our understanding about the regulation of camel CYP1A, 3A, and 2C gene expression, in relation to the up- or downregulation of appropriate promoter elements in camel cell lines. Moreover, it is important to analyze the substrate specificity of these proteins in greater detail to obtain an appropriate level of functional characterization. Clearly, much effort is required to identify and to characterize the isolated *C. dromedarius* CYPs, and to identify more camel CYPs within the same gene family.

ACKNOWLEDGMENTS

Conflicts of interest

The authors declare that there is no conflict of interest for this article and there is no financial employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, royalties related to this manuscript.

REFERENCES

Al-Khedhairy AA (2004). Characterization of the nucleotide sequence of a polyubiquitin gene (PUBC1) from Arabian camel, *Camelus dromedarius*. *J. Biochem. Mol. Biol.* 37: 144-147.

Alanazi MS, Saeed HM, Ataya FS and Bazzi MD (2010). Molecular characterization of the *Camelus dromedarius* putative cytochrome P450s genes. *Protein J.* 29: 306-313.

Andersson H, Kappeler F and Hauri HP (1999). Protein targeting to endoplasmic reticulum by dilysine signals involves direct retention in addition to retrieval. *J. Biol. Chem.* 274: 15080-15084.

Bar-Nun S, Kreibich G, Adesnik M, Alterman L, et al. (1980). Synthesis and insertion of cytochrome P-450 into endoplasmic reticulum membranes. *Proc. Natl. Acad. Sci. U. S. A.* 77: 965-969.

Bradford MM (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 72: 248-254.

el Sheikh HA, Ali BH, Homeida AM and Hapke HJ (1991). Activities of glutathione-*S*-transferase and ethoxycoumarin-*O*-deethylase in tissues of camels, sheep, goats and rats. *Comp. Biochem. Physiol. C* 98: 293-297.

Evans WE and Relling MV (1999). Pharmacogenomics: translating functional genomics into rational therapeutics. *Science* 286: 487-491.

Gonzalez FJ and Gelboin HV (1994). Role of human cytochromes P450 in the metabolic activation of chemical carcinogens

and toxins. *Drug Metab. Rev.* 26: 165-183.

Guengerich FP (1997). Comparisons of catalytic selectivity of cytochrome P450 subfamily enzymes from different species. *Chem. Biol. Interact.* 106: 161-182.

Guengerich FP (2004). Cytochrome P450: what have we learned and what are the future issues? *Drug Metab. Rev.* 36: 159-197.

Hasemann CA, Kurumbail RG, Boddupalli SS, Peterson JA, et al. (1995). Structure and function of cytochromes P450: a comparative analysis of three crystal structures. *Structure* 3: 41-62.

Hasler JA, Estabrook R, Murray M and Pikuleva I (1999). Human cytochromes P450. *Mol. Asp. Med.* 20: 1-137.

Hussein MF, Al-Moen AK and Gader AMA (1992). Haemostatic parameters in the camel (*Camelus dromedarius*) comparison with human. *Comp. Hematol. Int.* 2: 92-96.

Jung ST, Lauchli R and Arnold FH (2011). Cytochrome P450: taming a wild type enzyme. *Curr. Opin. Biotechnol.* 22: 809-817.

Kassner RJ (1973). A theoretical model for the effects of local nonpolar heme environments on the redox potentials in cytochromes. *J. Am. Chem. Soc.* 95: 2674-2677.

Kirby GM, Wolf CR, Neal GE, Judah DJ, et al. (1993). *In vitro* metabolism of aflatoxin B1 by normal and tumorous liver tissue from Thailand. *Carcinogenesis* 14: 2613-2620.

Laemmli UK (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227: 680-685.

Michael M and Doherty MM (2005). Tumoral drug metabolism: overview and its implications for cancer therapy. *J. Clin. Oncol.* 23: 205-229.

Mimura J and Fujii-Kuriyama Y (2003). Functional role of AhR in the expression of toxic effects by TCDD. *Biochim. Biophys. Acta* 1619: 263-268.

Nebert DW and Dalton TP (2006). The role of cytochrome P450 enzymes in endogenous signalling pathways and environmental carcinogenesis. *Nat. Rev. Cancer* 6: 947-960.

Nelson DR (2011). Progress in tracing the evolutionary paths of cytochrome P450. *Biochim. Biophys. Acta* 1814: 14-18.

Nelson DR and Strobel HW (1987). Evolution of cytochrome P-450 proteins. *Mol. Biol. Evol.* 4: 572-593.

Nelson DR, Kamataki T, Waxman DJ, Guengerich FP, et al. (1993). The P450 superfamily: update on new sequences, gene mapping, accession numbers, early trivial names of enzymes, and nomenclature. *DNA Cell Biol.* 12: 1-51.

Nelson DR, Zeldin DC, Hoffman SM, Maltais LJ, et al. (2004). Comparison of cytochrome P450 (CYP) genes from the mouse and human genomes, including nomenclature recommendations for genes, pseudogenes and alternative-splice variants. *Pharmacogenetics* 14: 1-18.

Ortiz AR, Strauss CE and Olmea O (2002). MAMMOTH (matching molecular models obtained from theory): an automated method for model comparison. *Protein Sci.* 11: 2606-2621.

Raza H, John A, Lakhani MS, Ahmed I, et al. (1998). Multiplicity and tissue specific expression of camel cytochrome P450(s). *Comp. Biochem. Physiol. C Pharmacol. Toxicol. Endocrinol.* 121: 205-211.

Raza H, Bhagwat SV and John A (2004). Flavin-containing monooxygenase activity in camel tissues: comparison with rat and human liver enzymes. *Comp. Biochem. Physiol. C Toxicol. Pharmacol.* 139: 289-293.

Rooney PH, Telfer C, McFadyen MC, Melvin WT, et al. (2004). The role of cytochrome P450 in cytotoxic bioactivation: future therapeutic directions. *Curr. Cancer Drug Targets.* 4: 257-265.

Roy A, Kucukural A and Zhang Y (2010). I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protocols* 5: 725-738.

Sakaguchi M, Mihara K and Sato R (1984). Signal recognition particle is required for co-translational insertion of cytochrome P-450 into microsomal membranes. *Proc. Natl. Acad. Sci. U. S. A.* 81: 3361-3364.

Sambrook J, Frisch E and Maniatis T (1989). Molecular Cloning: A Laboratory Manual. 2nd edn. Cold Spring Harbor Laboratory Press, New York.

Sanger F, Nicklen S and Coulson AR (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* 74: 5463-5467.

Seliskar M and Rozman D (2007). Mammalian cytochromes P450--importance of tissue specificity. *Biochim. Biophys. Acta* 1770: 458-466.

Shah PP, Saurabh K, Pant MC, Mathur N, et al. (2009). Evidence for increased cytochrome P450 1A1 expression in blood lymphocytes of lung cancer patients. *Mutat. Res.* 670: 74-78.

Stiborova M, Rupertova M and Frei E (2011). Cytochrome P450- and peroxidase-mediated oxidation of anticancer alkaloid ellipticine dictates its anti-tumor efficiency. *Biochim. Biophys. Acta* 1814: 175-185.

Thomas D, Bunch Warren C and Foote AM (1985). Chromosome banding pattern homologies and NORs for the Bacterian camel, guanaco and IIama. *J. Hered.* 76: 115-118.

Towbin H, Staehelin T and Gordon J (1979). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose

sheets: procedure and some applications. *Proc. Natl. Acad. Sci. U. S. A.* 76: 4350-4354.

Uppstad H, Ovrebo S, Haugen A and Mollerup S (2010). Importance of CYP1A1 and CYP1B1 in bioactivation of benzo[a] pyrene in human lung cell lines. *Toxicol. Lett.* 192: 221-228.

Yasukochi T, Okada O, Hara T, Sagara Y, et al. (1994). Putative functions of phenylalanine-350 of *Pseudomonas putida* cytochrome P-450cam. *Biochim. Biophys. Acta* 1204: 84-90.