HHS Public Access

Author manuscript

Mol Phylogenet Evol. Author manuscript; available in PMC 2017 January 01.

Published in final edited form as:

Mol Phylogenet Evol. 2016 January ; 94(Pt B): 676–687. doi:10.1016/j.ympev.2015.09.012.

Genetic and structural analyses of cytochrome P450 hydroxylases in sex hormone biosynthesis: Sequential origin and subsequent coevolution

Jared V. Goldstone[2,*], **Munirathinam Sundaramoorthy**[1,*], **Bin Zhao**[1], **Michael R. Waterman**[1], **John J. Stegeman**[2,#], and **David C. Lamb**[3,#]

[1]Department of Biochemistry, Vanderbilt University School of Medicine, Nashville, Tennessee, 37232-0146, USA

[2]Biology Department, Woods Hole Oceanographic Institution, Woods Hole, MA, 02543, USA

[3]Institute of Life Science, Medical School, Swansea University, Singleton Park, Swansea, SA2 8PP, UK

Abstract

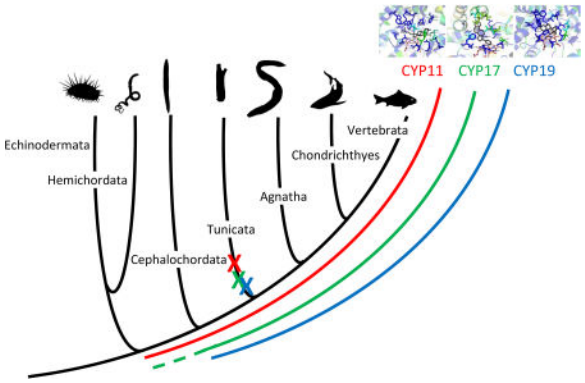
Biosynthesis of steroid hormones in vertebrates involves three cytochrome P450 hydroxylases, CYP11A1, CYP17A1 and CYP19A1, which catalyze sequential steps in steroidogenesis. These enzymes are conserved in the vertebrates, but their origin and existence in other chordate subphyla (*Tunicata* and *Cephalochordata*) have not been clearly established. In this study, selected protein sequences of CYP11A1, CYP17A1 and CYP19A1 were compiled and analyzed using multiple sequence alignment and phylogenetic analysis. Our analyses show that cephalochordates have sequences orthologous to vertebrate CYP11A1, CYP17A1 or CYP19A1, and that echinoderms and hemichordates possess CYP11-like but not CYP19 genes. While the cephalochordate sequences have low identity with the vertebrate sequences, reflecting evolutionary distance, the data show apparent origin of CYP11 prior to the evolution of CYP19 and possibly CYP17, thus indicating a sequential origin of these functionally related steroidogenic CYPs. Co-occurrence of the three CYPs in early chordates suggests that the three genes may have coevolved thereafter, and that functional conservation should be reflected in functionally important residues in the proteins. CYP19A1 has the largest number of conserved residues while CYP11A1 sequences are less conserved. Structural analyses of human CYP11A1, CYP17A1 and CYP19A1 show that critical substrate binding site residues are highly conserved in each enzyme family. The results emphasize that the steroidogenic pathways producing glucocorticoids and reproductive steroids are several hundred million years old and that the catalytic structural elements of the enzymes have been

[#]Corresponding authors: Drs. John Stegeman, Biology Department, Woods Hole Oceanographic Institution, Woods Hole, MA, 02543, USA, jstegeman@whoi.edu, and David C. Lamb, Institute of Life Science, Medical School, Swansea University, Singleton Park, Swansea, SA2 8PP, UK D.C.Lamb@swansea.ac.uk.

[*]Equal contribution

conserved over the same period of time. Analysis of these elements may help to identify when precursor functions linked to these enzymes first arose.

Graphical abstract



Keywords

cytochrome P450; steroid biosynthesis; pregnenolone; estrogen; progesterone; enzyme evolution

1. Introduction

Steroid hormones play important reproductive and developmental roles in vertebrates. Biosynthesis of steroid hormones involves enzymes belonging to two classes, the cytochrome P450 monooxygenases (CYPs) and the hydroxysteroid dehydrogenases (HSDs), both of which have been studied extensively (Miller et al., 2008; Payne and Hales, 2004). There are three CYP enzymes, the cholesterol side chain cleavage enzyme (CYP11A1), the bifunctional 17α-hydroxylase/17,20-lyase (CYP17A1), and the aromatase (CYP19A1), and two hydroxysteroid dehydrogenases (3β-HSD and 17β-HSD), in the steroid hormone biosynthetic pathways seen in Figure 1. CYPs are heme-containing monooxygenases most often catalyzing the conversion of a substrate (S) to a more polar hydroxylated form (S-OH). They bind O_2 much as hemoglobin and myoglobin do but reduce the molecular oxygen using electrons from pyridine nucleotides (NADH/NADPH) supplied through the ancillary redox partner proteins. *CYP*s constitute one of the largest gene superfamilies, and are found in all biological kingdoms. The evolution of different *CYP* families in the superfamily has been described (Nelson et al., 2013), however, the functional evolution of related *CYPs* has seldom been considered. Herein we present a structural analysis and evaluate the conservation of key features of these three CYP proteins.

The three steroidogenic CYPs each catalyze complex multi-step chemistry in conversion of substrate to product (Figure 1). The side chain cleavage enzyme, CYP11A1, catalyzes the conversion of cholesterol into pregnenolone (PREG) in three successive monooxygenation reactions, 22-hydroxylation, 20-hydroxylation and C22-C20 bond cleavage (Pikuleva, 2006; Storbeck et al., 2007). There are ohnologs of CYP11A1 (CYP11B1 and CYP11C1) that catalyze different reactions, but CYP11A1 catalyzes the first step in adrenal and gonadal steroidogenesis and is considered to be the rate-limiting step in steroid hormone

biosynthesis. Each of the three steps required for the conversion of cholesterol to PREG uses one molecule of both NADPH and O_2. CYP11A1 is a nuclear gene encoding a mitochondrial P450, and its mitochondrial electron transfer system consists of the 2Fe-2S protein adrenodoxin (Adx) and the FAD-containing reductase, adrenodoxin reductase (AdxR). The crystal structure of bovine CYP11A1 bound to the first reaction intermediate 22*R*-hydroxycholesterol and human CYP11A1 bound to the second reaction intermediate 20α,22*R*-dihydroxycholesterol have been reported from two different studies (Mast et al., 2011; Strushkevich et al., 2011).

CYP17A1 catalyzes two different reactions in the production of the androgens dehydroepiandrosterone (DHEA) and androstenedione (AND) (Miller, 2002; Nakajin et al., 1981). The production of DHEA requires only CYP17A1 whereas the production of AND additionally requires 3β-HSD. PREG produced by CYP11A1 serves as the substrate for CYP17A1, which first converts it into 17OH-pregnenolone (17OH-PREG) in a standard hydroxylation reaction and then produces DHEA by a 17,20-lyase reaction, a two-step process leading to cleavage of acetic acid from 17OH-PREG. The other androgen AND is produced from PREG or progesterone (PRG) by a combination of 17α-hydroxylation and 17,20-lyase reactions catalyzed by CYP17A1 and the 3β-dehydrogenation reaction catalyzed by 3β-HSD. Either PREG, 17OH-PREG or DHEA can act as substrate for 3β-HSD and the order in which these three reactions occur does not appear to be important. Finally, testosterone (TEST) is produced either directly from AND by 17β-HSD or from DHEA by 3β-HSD and 17β-HSD in two sequential reactions. CYP17A1 is a microsomal P450 and requires the microsomal NADPH-containing cytochrome P450 reductase (POR), which donates two electrons for the activation of molecular oxygen. Similar to CYP17A1, 3β-HSD and 17β-HSD are also microsomal enzymes. Crystal structures of human CYP17A1 in complex with inhibitors and substrates have been reported (DeVore and Scott, 2012; Petrunak et al., 2014).

Aromatase (CYP19A1) is unique in that it is the only enzyme in vertebrates that is capable of aromatization of a six-membered ring. CYP19A1 is involved in the production of three different estrogens, estrone, 17α-estradiol, and 17β,16α-estriol from the androgen substrates AND, Test, and 16αOH-Test, respectively. CYP19A1 is also a microsomal P450 requiring POR as the electron-transfer partner. The crystal structure of human CYP19A1 has been reported (Ghosh et al., 2009, 2010) and the analysis of CYP19A1 sequences from a number of species shows conservation of substrate binding residues (Hong et al., 2009, 2010).

Steroid hormones are found widely in animals. However, there are major differences in the structures and biosynthetic pathways among different classes of steroids in vertebrates (*Deuterostomata*), and insects and nematodes (*Ecdysozoa*). Markov et al. (Markov et al., 2009) assessed possible origins of enzymes including CYPs in the steroid hormone synthetic pathways in vertebrates by screening metazoan genomes using phylogenetic methods. They concluded that xenobiotic metabolizing CYPs are ancestors of steroidogenic CYPs, while steroidogenic CYPs of vertebrates and ecdysozoans arose in independent paths. Further, Markov et al (Markov et al., 2009) suggest that two steroidogenic CYPs, CYP11A1 and CYP19A1, arose in the deuterostome line and are specific to vertebrates. The only invertebrate shown to date to contain a CYP19 ortholog is the amphioxus *Branchiostoma*

belcheri (Mizuta and Kubokawa, 2007), an early-diverging chordate. Markov et al (Markov et al., 2009) also suggested that CYPs related to CYP11A1 and CYP17A1, cloned from amphioxus, are not orthologs of vertebrate CYP11A1 and CYP17A1 (Markov et al., 2009). Analysis of syntenic relationships suggested that the CYP19 gene may have arisen much earlier and been lost in many groups (Nelson et al., 2013). That same analysis (Nelson et al., 2013) also countered the argument (Markov et al., 2009) that CYP11 is strictly vertebrate, or even strictly chordate (Baker et al., 2014). There remains some uncertainty as to the phylogenetic origin of the steroidogenic CYPs.

Thus, in the present study, we have analyzed the structural information for the three CYP enzymes of vertebrate steroidogenic pathways, CYP11A1, CYP17A1 and CYP19A1 (Figure 1). A compilation of protein sequences and crystal structures of these three CYPs were examined for structural conservation related to function, in the context of complementary evolutionary studies of these multifunctional enzymes.

2. Materials and Methods

2.1 Sequence datasets

Protein sequences of CYP19, CYP17, and CYP11 were retrieved from the ENSEMBL, NCBI, and JGI protein sequence databases. The ENSEMBL database was used to obtain sequences of all three genes from each representative species (Supplemental Table 1), in order to provide consistent representation across all gene families. The amphioxus genome was queried at the Joint Genome Institute (JGI) database. Additional genomes searched include the echinoderms *Lytechinus variegatus* and *Patiria miniata*, the tunicates *Botryllus scholosseri* and, and the acorn worm *Saccoglossus kovalevskii*, (Supplemental Table 2). Genomes were searched using translated BLAST to identify nucleotide sequences with similarity to vertebrate CYP11, CYP17, or CYP19. Protein prediction (Genscan (Burge and Karlin, 1997), GenomeScan (Yeh et al., 2001), or FGeneSH (Solovyev et al., 2006)) was performed as needed to refined protein predictions. Homology was assessed using reciprocal best BLAST prior to phylogenetic analysis.

The NCBI database was used to obtain as many sequences as possible for each CYP family, using BLAST searching to identify vertebrate orthologs. Protein sequences were aligned using either ClustalW using the optimized alignment parameters of Hall (1999), or M-Coffee (Moretti et al., 2007). Alignments were visualized in BioEdit (Hall, 1999). Alignments used in the phylogenetic analyses are included as fasta files in the supplementary data.

2.2 Phylogenetic analyses

Phylogenetic trees were estimated by maximum likelihood (ML) using RAxML (v8.0.1) with the PROTWAGCAT and PROTLGCAT model of amino acid substitution (Stamatakis, 2014)Replicate ML trees were constructed, and the best-scoring tree was subjected to rapid bootstrapping in RAxML to assess bifurcation support using the extended majority rule (MRE) bootstopping criterion.

2.3 Homology assignment

Phylogenetic methods, the active site residues, and (where possible) synteny were used to assign CYP family membership. A CYP family is defined by a 40% cutoff in overall amino acid identity, a definition that was created in the late 1980s with limited data availability. This definition is remarkably consistent with functional classification within vertebrates, yet is not concordant with phylogenetic analyses across broader taxonomic divisions (Nelson et al., 2013). Thus, we have relied more on phylogenetic and active site residue analyses to guide our assignments. Reciprocal best blast hits were used to analyze and filter genome-wide searches prior to phylogenetic analysis.

2.4 Structural Analyses

Coordinates of human CYP19A1 (PDB: 3EQM, human CYP17A1 (PCB: 3RUK) and human CYP11A1 (PDB: 3N9Y) were analyzed using PyMOL [The PyMOL Molecular Graphics System, Version 1.6, Schrödinger, LLC.] Homology models of zebrafish CYP19A1, CYP19A2, CYP17A1, CYP17A2, CYP11A1, and CYP11A2 were constructed using Modeller (v9.11) (Webb and Sali, 2014) based on the human structures. Homology modeling was carried out by satisfaction of spatial restraints using the automodel function of Modeller, with very thorough VTFM, thorough MD, and two repeat cycles of minimization. The best model from the generated structures was selected based on the Discrete Optimized Protein Energy (DOPE) score (Eramian et al., 2006; Shen and Sali, 2006), and further assesed using Procheck (Laskowski et al., 1996).

3. Results

3.1 Origins of key steroidogenic CYP genes

The existence of steroidogenic pathways has been well established for vertebrates, and extended to members of the clade *Craniata*. The presence of these pathways in early deuterostomes is less clear. The cephalochordate *B. belcheri* (amphioxus, or lancelet) has been shown to contain sex steroids in the gonads, and cloning of *CYP* genes suggested to encode CYP11A, CYP17A, and CYP19A from the ovaries of amphioxus (Mizuta and Kubokawa, 2007) implies that cephalochordates may have a steroid hormone biosynthetic pathway similar to the vertebrate pathway. A partial sequence of CYP19A cloned from another lancelet species *Branchiostoma floridae* also shows a strong similarity to the vertebrate CYP19A1 sequences (Castro et al., 2005; Mizuta and Kubokawa, 2007). This conclusion is supported by a recent phylogenetic analysis of the vertebrate steroidogenic CYPs (Baker et al., 2014).

Tunicates are the closest living deuterostome relatives of vertebrates (Delsuc et al., 2006), but there is no evidence for the presence of a complete sex hormone synthesis pathway in the subphylum *Tunicata* (Campbell et al., 2004). In fact, the genomes of the ascidian tunicate species *Ciona intestinalis* and *Ciona savignyi* lack CYP51 (lanosterol demethylase), the enzyme required for the biosynthesis of the steroid precursor cholesterol (Morrison et al., 2014). CYP51 also appears to be missing from the genomes of the appendicularian *Oikopleura dioica* and the molgulid *Molgula occidentalis*, suggesting that CYP51 was lost early in tunicate evolution. However, the sex steroids estradiol and testosterone and their

precursors and metabolites have been reported in *C. intestinalis* (Cangialosi et al., 2010). The *C. intestinalis* genome has been reported to have orthologs of the initial steroidogenic enzymes CYP11A1 and 3β-HSD, but lacks the other key enzymes CYP17A1 and CYP19A1 (Campbell et al., 2004).

We used the CYP11A1, CYP17A1, and CYP19A1 sequences of human and *B. belcheri* to search databases for similar sequences. There are 63 species for which partial or complete versions of all three genes are available in the databases (Supplemental Table S1). Two of the three sequences were available for several other species, while for some only one was found (Table S1), including partial sequences.

Search of the NCBI database using the reported *B. belcheri* CYP11 (BAF61103.1; denoted CYP11_BB1) sequence [12] identified two related sequences from *B. floridae* (XP_002604199.1 termed CYP11_BF1, and XP_002611984.1 termed CYP11_BF2), one sequence from the sea urchin *Strongylocentrotus purpuratus* (XP_781842.1, termed CYP11-SP1), and sequences from the hemichordate acorn worm *Saccoglossus kowalevskii (*Sakowv30013852m and Sakowv30009358m*)*. Several mitochondrial CYP sequences were identified from *S. kowalevskii*, however phylogenetic analysis indicates that only two are more closely related to CYP11 than to CYP27 (Figure 2A).

While the 526-residue long CYP11_BF2 of *B. floridae* is comparable in length to the *B. belcheri* CYP11 sequence, the 882 residue long CYP11_BF1 is interspersed with long non-P450 sequence fragments. These sequences are different from the 18 transcripts that are annotated as CYPs belonging to the mitochondrial clan in the *B. floridae* genome. This discrepancy prompted us to compare the 18 putative mitochondrial CYP sequences with a consensus sequence built using CYP11A1 sequences from the vertebrate species identified by searching with the human CYP11A1 sequence. Pair-wise alignment of these *B. floridae* sequences with the consensus CYP11A1 sequence identified a full-length sequence (XP_002611060.1, termed CYP11_BF3) and a partial sequence (XP_002607853.1, termed CYP11_BF4) with significantly greater identities to CYP11A1 (33.3% and 28.7%, respectively, for the full length proteins) than the remaining 16 sequences. Due to the fragmentary nature of the *B. floridae* CYP11_BF1 and BF4 gene predictions, and the high degree of similarity (more than 60% identity over 361 positions for CYP11_BF1) with the known *B. belcheri* CYP11 sequence, the *B. floridae* CYP11_BF1 and CYP11_BF4 were not included in phylogenetic analyses. Due to the low sequence identity with CYP11As, in a recent publication CYP11_BF1 and CYP11_BF4 have been named CYP11D, while the CYP11_BF2 and CYP11_BF3 have been named CYP11Es (Baker et al., 2014).

Searches using CYP17 also retrieved two sequences from *B. floridae* (XP_002612045.1, Bfloridae_17BF1; and XP_002601863.1, Bfloridae_17BF2), one sequence each from the sea urchins *S. purpuratus* (XP_789963.1, Spurpuratus_17) and *Mesocentrotus nudus* (ADA81873.1, Mnudus_17), and a sequence from the acorn worm *S. kowalevskii* (XP_002730849.1, Skowalevskii _17). The Bfloridae17_BF1 sequence shows significantly greater identity (40%) with the human CYP17A1 sequence than does the Bfloridae17_BF2 sequence (29.8%), and thus the Bfloridae17_BF1 should be named CYP17C1 (as lamprey have a CYP17 divergent enough to be termed CYP17B1 (Baker et al., 2014)). The trend is

reversed when these *B. floridae* sequences are compared with the *B. belcheri* sequence (33.3% for CYP17_BF1 and 80% for CYP17_BF2). Functional properties of these sea urchin and worm CYPs have not been determined and thus assignment as either 17α-hydroxylase or 17,20-lyase function is inappropriate at present.

No CYP19 homolog was detected in either the echinoderms or the acorn worm. As noted above, however, CYP19 has been cloned from *B. floridae* (BAF61105.1) (Castro et al., 2005; Mizuta and Kubokawa, 2007).

3.2 Phylogenetic analyses

As is the case for many proteins, it is extremely difficult to point out one or two regions that clearly distinguish hypothesized functional similarities, and thus analysis of the active sites and full length protein sequences are important to assigning orthology. Separate phylogenetic trees were constructed for alignments of the inferred amino acid sequences for CYP11s, CYP17s and CYP19s from selected species (Table 1) for which all members of the three CYP families were available, and including the proposed homologous CYPs from selected early-diverging deuterostomes (as discussed above) (Figure 2A, 2B, 2C). Additional phylogenetic trees were constructed for a given *CYP* family using all sequences available (Supplemental material, Figures S1, S2, and S3). As found previously, the teleost fish *CYP19* genes cluster in two separate clades, the *CYP19A1s* and *CYP19A2s* (Wilson et al., 2005). Human CYP24A1 and CYP27A1 were used as outgroup sequences in the CYP11 phylogeny. The two *CYP11A* genes in zebrafish cluster together, which is a reasonable result as these genes likely represent a recent, tandem duplication (Goldstone et al., 2010); similarly the shark *C. milli* CYP11A1 and CYP11A2 are together. Notably, also we found a cluster of invertebrate sequences including two of the amphioxus CYP11 sequences (Bflo_CYP11_BF2 and the Bbel sequence) at the base of the vertebrate CYP11A clade. Tetrapod CYP11Bs (11-beta hydroxylase and aldosterone synthase) appear to have evolved just before or after the transition to land (Colombo et al., 2006), but, although present in alligators, are missing in birds.

There are two *CYP17* genes in fish, encoding CYP17A1 and CYP17A2, including in the elephant shark (*Callorhinchus milii*), the spotted gar (*Lepisosteus oculatus*) and the coelacanth (*Latimeria chalumnae*). These three species were not subjected to the teleost whole genome duplication (WGD3), and thus the CYP17A duplication occurred between cephalochordate and elasmobranch diversification. The CYP17A1 and CYP17A2 clusters fall into separate clades when using the amphioxus genes as the root. Notably, the tetrapod *CYP17A1* genes cluster together with the fish *CYP17A1* genes, indicating that *CYP17A2* was lost from the tetrapod lineage, a conclusion reached recently by others (Baker et al., 2014; Nelson et al., 2013). The amphioxus CYP17 sequences cluster with the vertebrate CYP17s. Two of the *S. purpuratus* genes also cluster with vertebrate CYP17s, as does a S. kowalevskii gene, although with non-significant bootstrap support. Other *S. purpuratus* and *M. nudus* CYP17-like genes do not (the phylogeny is rooted with human CYP1A1), and indeed most of these genes also branch outside of the vertebrate CYP21As (represented by human CYP21A2).

3.3 Sequence Conservation and Structure Analyses

The alignment of all CYP11 sequences, including the reported *B. belcheri* sequence and the predicted *B. floridae* sequences, shows 34 identical residues, 91 strongly conserved residues and 72 weakly conserved residues, respectively (Figure 3, Supplementary Fasta File 1). The *B. belcheri* sequence shows the greatest difference from the rest of the sequences, as it is only 26% identical to the human sequence (however, if ambiguously aligned residues are removed, this identity rises to 33%). Similarly, the *B. floridae* sequences also display low amino acid identities to the human CYP11A1. If the *B. belcheri* sequence is excluded, the alignment of the remaining 36 sequences improves significantly showing 64 identical residues, and 79 strongly conserved and 33 weakly conserved, respectively.

Most of the conserved residues in CYP11A1 may play either structural or functional roles. There are several conserved glycine and proline residues that probably play structural roles in protein folding, serving as helix breakers or terminators. Other conserved residues likely play important roles in substrate access to or binding in the active site, or for the interaction of CYP11A1 with the electron transfer partner Adx.

A previous site-directed mutagenesis study identified two highly conserved lysine residues (Lys377 and Lys381) in bovine CYP11A1 that are essential for Adx binding (Wada and Waterman, 1992). The first lysine is conserved in all but *Monodelphis domestica* CYP11A1, and the second lysine is absolutely conserved in all the sequences examined in the present analysis. Not only are these lysines highly conserved in CYP11A1 sequences, they are also conserved in all six confirmed mitochondrial CYPs in the human genome, suggesting that the binding of Adx to mitochondrial CYPs is highly specific. An exhaustive analysis of all known mitochondrial CYP sequences, which is beyond the scope of the present study, might throw more light on this prediction. However, this hypothesis does derive support from the analysis of Adx sequences from nearly 20 species (data not shown). Two aspartic acid residues, Asp76 and Asp79, in human Adx were shown to be important for binding to CYP11A1 and these two residues are among several absolutely conserved residues in Adx sequences (Coghlan and Vickery, 1991). These observations strongly suggest that a mitochondrial CYP and its electron transfer partner, Adx, interact with high specificity through electrostatic interactions of oppositely charged residues in the two proteins.

Several crystal structures of CYP11A1 with bound steroid substrates have been solved (Mast et al., 2011; Strushkevich et al., 2011). Analysis of the human structure reveals that the substrate is surrounded by residues Phe121, Ile123, Trp126, Leu140, Met240, Phe241, Leu248, Glu322, Ala325, Ile390, Ser391, Val392, Thr393, Gln395, Gln416, and Leu499 (Figure 4C). Many of these residues are conserved in most species whose sequences are aligned in our study. Residues Trp126, Met240, Phe241, Leu248, Glu322, Ala325, Gln416, and Leu499 are absolutely conserved among mammals but have a few minor differences in the amphioxus, hemichordate and echinoderm sequences. Similarly, Leu140, Gln395, and Gln416 are highly conserved within vertebrates, but are different in the invertebrate CYP11 sequences (e.g. L140 is conserved in acorn worms, but is substituted with a valine or an isoleucine in amphioxus and sea urchin, but Q416 is not conserved at all in the invertebrate CYP11s). Also, a few residues are substituted by similar amino acids; thus, residue 390 is

either Ile or Val in all sequences, and 392 is either Val or Ile in most sequences (but a Phe in most invertebrates). Only two active site residues are not highly conserved. Phe121 is not conserved within vertebrate, being substituted by Tyr, Leu, Ile or Met residues, and Ser391 is substituted with Ala in many sequences. Notably, the more distantly related invertebrate deuterostomes CYP11-like sequences also show a high degree of sequence conservation in these substrate binding and positioning residues (Table 1). This supports the phylogenetic analysis, and indicates that these latter CYPs are likely functionally conserved as well.

Alignment of selected vertebrate CYP17A sequences shows 53 identical, 92 strongly similar and 36 weakly conserved residues (Figure 3, Supplementary Fasta File 2). The addition of the more distantly related CYP17s from *B. floridae* (CYP17_BF2), and from sea urchin and acorn worm substantially reduce the number of identical and strongly similar sequences.

Analysis of a three-dimensional structure of human CYP17A1 (PDB: 3RUK (DeVore and Scott, 2012)) shows residues Ala105, S106, Phe114, Tyr201, Asn202, Ile206, Leu209, Gly297, Asp298, Ala302, Thr306, V366, Ile371 and Val483 surround the abiraterone molecule in the active site (Figure 4B) and most of these residues are hydrophobic. Most of these residues are highly conserved among the sequences aligned in this study (Figure 4B). The residue Phe114 is absolutely conserved in all sequences and also residues Asp298, Ala302, and Thr306 are highly conserved. The bound ligand abiraterone is an inhibitor, and the mode of binding of the substrate(s) can be different from that of inhibitors, although a recent analysis of human CYP17A1 co-crystallized with several substrates did not highlight such differences (Petrunak et al., 2014). That analysis did emphasize the importance of Ala105 and Asn202 in active site orientation of native substrates.

A lesser degree of identity between the CYP17A1 sequences of early diverging deuterostomes such as *B. flordiae* and later-diverging deuterostomes (mammals) reflects evolutionary distance. However, the lesser identity between bifunctional and monofunctional teleost CYP17A1 and 17A2 sequences could suggest residues that distinguish the lyase and the 17-hydroxylase activities. Thus, the multiple sequence alignment including seven fish CYP17A2 sequences reveals that nine fewer residues are absolutely conserved than when those seven sequences are excluded from the alignment. In human CYP17A1, these additionally conserved residues correspond to Gln57, Ala82, Ile146, Cys183, Asp298, Ile357, and Leu361 (Figure 4B). These residues might be used to identify the 17,20-lyase activity in CYP17A1 by mutagenesis studies. However, recent crystal structures of zebrafish CYP17A1 and CYP17A2 indicate that substrate activity differences between the zebrafish isoforms are not easily ascribable to individual active site amino acid differences, and further work is necessary even though good structural information is available (Gilep et al., 2011; Pallan et al., 2015).

The larger CYP19 sequence dataset comprises 283 complete sequences of which 165 are from fish species [BLAST cutoff 1e-27]. As expected, the alignment of all 283 sequences shows the least conservation in *N*- and *C*-terminal regions (Figure 3; Supplementary Fasta File 3). The fish ovary CYP19A1 sequences are generally longer, with additions at the *N*-terminus, compared to sequences of the fish brain CYP19A2 or those of other CYP19A1s. In contrast, the fish brain CYP19A2 sequences are characterized by a short polyleucine

motif at the *N*-terminus. Except for these differences at the *N*-terminus, the CYP19 sequences show a remarkably high degree of similarity as compared to the CYP11 and CYP17 sequence alignments (discussed later).

Inspection of the aligned CYP19 sequences reveals 50 identical, 76 strongly similar, and 28 weakly similar residues (Supplemental Fasta File 3). The crystal structure of human aromatase with the bound substrate AND shows that the side chains of residues Arg115, Ile133, Phe134, Phe221, Trp224, Ala306, Thr310, Val 370, Val 373, Met 374 and Leu477 make direct van der Waals contacts with the substrate [7] (Figure 4A). All these residues except Ala306, Val373 and Met374 are absolutely conserved in all sequences. There is only one exception for Ala306 (Ser in CYP19A2 of the fish *Halichoeres tenuis*) and for Met374 (Tyr in CYP19A1 of the fish *Ictalurus puntatus*). The residue Val373 has broader substitution with Ile (birds and CYP19A2 of fishes), Thr (fish CYP19A1 and CYP19A2), or Ser (CYP19A2 of fishes), while Val is conserved in mammals. These observations are in line with those of Callard et al (2011), who also described the strong sequence conservation between the amphioxus CYP19 and vertebrate CYP19s in the I helix, aromatic region, and heme binding domain (Callard et al., 2011).

4. Discussion

CYP11A1 is the first enzyme in the steroidogenesis pathway, and thus is essential for the synthesis of all three classes of steroid hormones (glucocorticoids, mineralocorticoids and sex steroids) (Figure 1). CYP17A1 is involved in the synthesis of glucocorticoids and sex hormones but is not required for the synthesis of mineralocorticoids, while CYP19A1 is required only for the synthesis of estrogens. Thus, concurrent with the hypothesis of the evolution of increasing systemic complexity (e.g. evolutionary ‘ratchets’ (Bridgham et al., 2009)), the evolution of complex steroid hormonal systems occurred most likely in a stepwise fashion. Accordingly, a CYP11 cholesterol side-chain cleavage enzyme necessarily evolved prior to a CYP19 aromatase that would work on the product of such a cleavage. Thus, rather than a coincident evolution for the steroidogenic CYPs in a chordate ancestor, we anticipate, and data support, a sequential origin. Although sequences of ancient common ancestral proteins are not available, sequences from modern descendent species, hemichordates and echinoderms, support the origin of CYP11 functionality prior to the evolution of CYP17, and CYP19. Subsequent coevolution was then constrained by the complexity of hormonal systems required for successful development and reproduction.

4.1 Steroid patterns in deuterostomes

Analysis of steroids in diverse species supports the conclusion of sequential steroidogenic enzyme evolution. PRG, E2, and TEST have all been detected in amphioxus ovary tissue, using radioimmune assays (Mizuta and Kubokawa, 2007). Additionally, *in vitro* conversion of radiolabeled steroidal enzyme substrates have been confirmed, although the measured enzyme activities were likely that of 3P/A5-4-HSD and 17P-HSD, rather than CYP11 or CYP17 (Mizuta et al., 2008). Interestingly, despite the lack of clear steroidogenic CYP orthologs in tunicates, several ovarian steroids have been identified in *Ciona intestinalis* (Cangialosi et al., 2010). However, as we recently noted, there is no lanosterol 14-a demethylase (CYP51) in either of the *Ciona* species’ genomes (Morrison et al., 2014).

CYP51 is required for de novo cholesterol synthesis leading to an obligate dietary or commensal source for steroid precursors in *Ciona* spp. (Morrison et al., 2014).

More clearly supporting the notion of sequential evolution of steroidogenesis steps are the observations of the conversion of cholesterol into PRG, and PRG to androgens, in the sea star *Asterias rubens* (Lafont and Mathieu, 2007; Schoenmakers and Voogt, 1981; Schoenmarkers and Voogt, 1980; Voogt et al., 1987). However, no estrogen aromatization has been found to occur in *A. rubens* or in *A. vulgaris*, consistent with the idea that CYP19 aromatase evolved after CYP11 and CYP17 (Hines et al., 1992; Schoenmakers and Voogt, 1981). Little work has been done on steroids in sea urchins. However, *S. purpuratus* appears to have a CYP11 gene, but to lack at least CYP19 (Goldstone et al., 2006). Almost nothing is known also about hemichordate steroids (there are only about 100 hemichordate species known). However a few C27 steroids were found in *S. kowalevskii* extracts, along with squalene intermediates (Carey and Farrington, 1989). Thus, it appears that the complex vertebrate steroidogenic pathway can be traced back through sequential gains of CYP enzyme activities over the course of deuterostome evolution.

4.3 Evolutionary considerations of vertebrate steroidogenic CYPs

Multiple genes for steroidogenic enzymes in teleost fish help to shed light on the evolution of these genes in vertebrates. Teleost fish species are known to have two isoforms of many genes, due to the ancient teleost-specific whole genome duplication event (“WG3”) estimated to have occurred approximately 350 million years ago (Meyer and Van de Peer, 2005). Thus, fish have two CYP19 isoforms, ovarian CYP19A1 (CYP19A1a) and brain CYP19A2 (or CYP19A1b) (Tchoudakova and Callard, 1998). All sequenced tetrapods appear to have only one gene encoding CYP19A1, with the exception of pig, which has three isoforms, from brain, placenta, and embryo (here termed CYP19A1, CYP19A3, and CYP19A4 (Corbin et al., 2003)).

Teleost fish species also have been shown to have two different *CYP17A* genes on separate chromosomes, encoding CYP17A1 and CYP17A2 proteins (Zhou et al., 2007a; Zhou et al., 2007b). Teleost CYP17A1 is a bifunctional enzyme exhibiting both 17α-hydroxylase and 17,20-lyase activities, while CYP17A2 is a monofunctional enzyme with 17α-hydroxylase activity but lacking 17,20-lyase activity (Zhou et al., 2007a; Zhou et al., 2007b). Although teleost CYP17A1 exhibits 17α-hydroxylase activity, they appear to be much more effective cholesterol-17,20-lyases (Pallan et al., 2015). No tetrapods are known to have two CYP17As, however coelacanth appears to have two expressed CYP17A genes (Amemiya et al., 2013).

Fish also have multiple CYP11A1 sequences, which in zebrafish are named CYP11A1 and CYP11A2. These are the result of a tandem duplication, rather than whole genome duplication, and exhibit high sequence identity, suggesting a relatively recent origin (Goldstone et al., 2010). The spotted gar (*Lepisosteus oculatus*) has only one CYP11A. In zebrafish, CYP11A1 is only expressed in early development in the gonads (until 48 hpf), while the CYP11A2 paralog is expressed in larvae and in adult steroidogenic tissues (Parajes et al., 2013). Thus CYP11A in zebrafish exhibits temporal subfunctionalization, rather than the functional differentiation like that seen for CYP17A1 and CYP17A2. Fish additionally

have CYP11C genes, and while CYP11C1 is known to be expressed in zebrafish testicular tissue and the cytoplasm of the follicular cells surrounding vitellogenic oocytes and of interstitial cells (Baudiffier et al., 2012; Caulier et al., 2015), the specific steroid metabolites of CYP11C1 are not known.

4.4 Origins of steroidogenic CYPs

B. belcheri is the only organism in the *Cephalochordata* for which all three steroidogenic CYP sequences have been reported (Castro et al., 2005; Mizuta and Kubokawa, 2007; Nelson et al., 2013). Based on BLAST searches we were able to identify probable CYP11, CYP17, and CYP19 genes also in *B. floridae*. The published sequence of the *B. belcheri* CYP11 and its apparent orthologs from *B. floridae* and *S. purpuratus* (Figure 2A) form a separate clade distinct from the vertebrate sequences, similar to the picture obtained for the CYP17s sequences from those species. Since the *B. belcheri* sequence and its orthologs from *B. floridae* (CYP11_BF1 and CYP11_BF2), are closer to *S. purpuratus* sequence than the vertebrate CYP11A1 sequences, there also are questions about the function of these invertebrate CYP11-like homologs.

The phylogenetic tree constructed with representative CYP17 sequences shows that the invertebrate sequences form a separate clade (Figure 2B). Within this clade, the *B. floridae* (CYP17_BF1) sequence does not cluster with other invertebrate sequences but rather groups with the vertebrate CYP17 sequences. As steroidogenic pathways are not known in echinoderms, whether the CYP17s in *Strongylocentrotus* spp. are functionally similar to vertebrate CYP17s is unclear. The same uncertainty exists for the reported *B. belcheri* sequence and *B. floridae_2* (CYP17_BF2). However, analysis of the sequence conservation around the active site supports the idea of a steroid substrate (see below). CYP17-like sequences have been observed in mollusks (Zanette et al., 2010), but functional or orthologous relationships to the deuterostome sequences are not clear. Phylogenetically, these mollusk sequences are slightly more divergent than the deuterostome sequences described here. Nevertheless, such sequences raise the possibility that a CYP17 homolog preceded the deuterostomes and has been altered or was lost in some lineages, including perhaps the Ecdysozoa. There is not good evidence for a CYP11-like sequence, however, although there are multiple (uncharacterized) mitochondrial CYPs (Zanette et al., 2010).

The CYP19 phylogenetic tree includes one sequence each from *B. belcheri* and *B. floridae* (Figure 2C), which have been identified in previous studies (Callard et al., 2011; Castro et al., 2005; Mizuta and Kubokawa, 2007). Even though these sequences appear relatively distant from those of the rest of the CYP19 sequences in the tree, they align well with the other CYP19s, and are distinct from the non-steroidogenic CYPs of other invertebrates. Since a vast majority of CYP19 sequences are from fish, only representative sequences from six different fishes have been included in the tree. A striking feature of the CYP19A1 phylogenetic tree is that the brain isoform CYP19A2 found in the fishes clusters with the CYP19A1s of other species; the fish ovarian CYP19A1s apparently diverged from the orthologous CYP19 line via the teleost whole genome duplication (Figure 2C).

Estrogen receptor (ER) and estrogen related receptor (ERR) sequences have been identified in mussels (*Mytilus* species) (Nagasawa et al., 2015), as well as in oyster (Vogeler et al.,

2014), and other molluscs (Keay et al., 2006; Thornton et al., 2003). This suggests there could be functional estrogen signaling in the *Lophotrochozoa*, however, molluscan steroid receptors (called ERs due to the high sequence similarity to vertebrate ERs) are not responsive to estrogen (Matsumoto et al., 2007) due to ancient amino acid substitutions rendering them unresponsive to steroids (Bridgham et al., 2014). This ancient evolution of an "allosteric switch" to constitutive activity is consistent then with the apparent absence of CYP19 in protostomes. Unfortunately, almost nothing is known in detail about the evolution and function of steroid nuclear receptors in echinoderms and acorn worms (Baker et al., 2015; Goldstone et al., 2006; Howard-Ashby et al., 2006).

4.5 Sequence Conservation and Structure Analyses

The functional divergence of vertebrate CYP19s has been examined, and that overall differences among them appear to be functionally neutral (Wilson et al., 2005). Hong and colleagues also analyzed a large number of CYP19 protein sequences including both brain and ovarian isoforms found in several fish species (Hong et al., 2009, 2010). The analysis of sequences and structural models identified conserved amino acid residues in CYP19A1 potentially important for binding of substrates and the electron transfer partner POR. An area less attended to are studies showing that aromatases undergo posttranslational modifications, both glycosylation and phosphorylation. Human, bovine, and equine aromatases are glycosylated (Sethumadhavan et al., 1991; Shimozawa et al., 1993), but not the porcine ovary and placenta isoforms (Corbin et al., 2003). The human aromatase sequence shows two potential N-glycosylation sites at Asn12 and Asn180 and it has been predicted that the glycosylation site is localized at the N-terminal membrane-spanning region (Shimozawa et al., 1993). Examination of the N-terminal region of the sequences in our data shows striking structural features that support these observations. Most mammalian aromatases including human, bovine and equine CYP19s have NXT/S (Asn12-Ile13-Thr14) motif at the N-terminal region, but porcine ovary and placenta sequences lack this motif. Interestingly, the porcine embryonic sequence contains NXT/S motif implying that this isoform is probably glycosylated, although this has not been tested experimentally.

The NXT/S motif is present only in mammalian and amphibian species but not in birds, reptiles or either of the CYP19 isoforms in fishes, although few exceptions are found. Studies using quail brain homogenates suggest that aromatase is phosphoregulated through a calcium dependent mechanism involving multiple kinases including protein kinases C (Balthazart et al., 2003; Balthazart and Ball, 2006). Site-directed mutagenesis of murine aromatase has demonstrated that Ser118 is a potential phosphorylation site and mutation of this site destabilized the enzyme and decreased the specific activity (Miller et al., 2008). Unlike the less conserved N-terminal glycosylation site, the predicted phosphorylation site Ser118 is conserved in all aromatase sequences analyzed in the present study except in the ovarian isoform of zebrafish (Figure 4C). The differences in the levels of conservation of glycosylation and phosphorylation correlate with their effect on the function of the enzyme. While the poorly conserved N-glycosylation site doesn't show any effect on the activity of aromatase (Corbin et al., 2003), the nearly absolutely conserved phosphorylation site has a significant effect on the enzyme activity, providing an additional kinase-mediated mechanism for modulating enzyme activity (Miller et al., 2008).

5. Conclusion

In conclusion, this study addresses the phylogeny of CYPs involved in key steps of reproductive and adrenal steroids, and details of sequence associated with function to assess the evolution of the pathways. The current phylogenetic analyses support the sequential origin of steroidogenic enzymes, though we recognize that additional genomes, for tunicates and other early deuterostomes, may resolve some uncertainties. Such uncertainties notwithstanding, the sequence alignment of the steroid hydroxylases described here identifies amino acids, which are conserved through different periods of deuterostome evolution. These active site amino acids are highly likely to be critical in the structure/function of these monooxygenases, and the active sites have been highly conserved once the three types of enzymes evolved in the *Chordata*. Together these results emphasize that the steroidogenic pathways producing glucocorticoids and sex hormones are several hundred million years old and that key structural elements involved in catalytic activity of the enzymes have existed over the same period of time. Continued searching with these key elements may provide clear answers as to the phylogenetic origin of CYP11, CYP17 and CYP19 functions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Support was provided by NIH U41HG003345 (JVG), NIH Superfund Research Program at Boston University P42ES007381 (JJS and JVG), and by a Wellcome Trust Sabbatical Grant to DCL.

References

Amemiya CT, et al. The African coelacanth genome provides insights into tetrapod evolution. Nature. 2013; 496:311–316. [PubMed: 23598338]

Baker ME, Nelson DR, Studer R. Origin of the response to adrenal and sex steroids: Roles of promiscuity and co-evolution of enzymes and steroid receptors. J Steroid Biochem Mol Biol. 2014 in press.

Baker ME, Nelson DR, Studer RA. Origin of the response to adrenal and sex steroids: Roles of promiscuity and co-evolution of enzymes and steroid receptors. J Steroid Biochem Mol Biol. 2015; 151:12–24. [PubMed: 25445914]

Balthazart J, Baillien M, Charlier TD, Ball GF. Calcium-dependent phosphorylation processes control brain aromatase in quail. Eur J Neurosci. 2003; 17:1591–1606. [PubMed: 12752377]

Balthazart J, Ball GF. Is brain estradiol a hormone or a neurotransmitter? Trends Neurosci. 2006; 29:241–249. [PubMed: 16580076]

Baudiffier D, Hinfray N, Vosges M, Creusot N, Chadili E, Porcher JM, Schulz RW, Brion F. A critical role of follicle-stimulating hormone (Fsh) in mediating the effect of clotrimazole on testicular steroidogenesis in adult zebrafish. Toxicology. 2012; 298:30–39. [PubMed: 22564764]

Bridgham JT, Ortlund EA, Thornton JW. An epistatic ratchet constrains the direction of glucocorticoid receptor evolution. Nature. 2009; 461:515–519. [PubMed: 19779450]

Bridgham JT, Keay J, Ortlund EA, Thornton JW. Vestigialization of an allosteric switch: genetic and structural mechanisms for the evolution of constitutive activity in a steroid hormone receptor. PLoS Genet. 2014; 10:e1004058. [PubMed: 24415950]

Burge C, Karlin S. Prediction of complete gene structures in human genomic. DNA J Mol Biol. 1997; 268:78–94. [PubMed: 9149143]

Callard GV, et al. Evolutionary origins of the estrogen signaling system: insights from amphioxus. J Steroid Biochem Mol Biol. 2011; 127:176–188. [PubMed: 21514383]

Campbell RK, Satoh N, Degnan BM. Piecing together evolution of the vertebrate endocrine system. Trends Genet. 2004; 20:359–366. [PubMed: 15262408]

Cangialosi MV, Puccia E, Mazzola A, Mansueto V, Arukwe A. Screening of ovarian steroidogenic pathway in *Ciona intestinalis* and its modulation after tributyltin exposure. Toxicol Appl Pharmacol. 2010; 245:124–133. [PubMed: 20214916]

Carey DA, Farrington JW. Polycyclic Aromatic-Hydrocarbons in Saccoglossus-Kowalewskyi (Agassiz). Estuarine Coastal and Shelf Science. 1989; 29:97–113.

Castro LF, Santos MM, Reis-Henriques MA. The genomic environment around the Aromatase gene: evolutionary insights. BMC Evol Biol. 2005; 5:43. [PubMed: 16098224]

Caulier M, Brion F, Chadili E, Turies C, Piccini B, Porcher JM, Guiguen Y, Hinfray N. Localization of steroidogenic enzymes and Foxl2a in the gonads of mature zebrafish (Danio rerio). Comp Biochem Physiol A Mol Integr Physiol. 2015; 188:96–106. [PubMed: 26099948]

Coghlan VM, Vickery LE. Site-specific mutations in human ferredoxin that affect binding to ferredoxin reductase and cytochrome P450scc. J Biol Chem. 1991; 266:18606–18612. [PubMed: 1917982]

Colombo L, Dalla Valle L, Fiore C, Armanini D, Belvedere P. Aldosterone and the conquest of land. J Endocrinol Invest. 2006; 29:373–379. [PubMed: 16699307]

Corbin CJ, Mapes SM, Lee YM, Conley AJ. Structural and functional differences among purified recombinant mammalian aromatases: glycosylation, N-terminal sequence and kinetic analysis of human, bovine and the porcine placental and gonadal isozymes. Mol Cell Endocrinol. 2003; 206:147–157. [PubMed: 12943997]

Delsuc F, Brinkmann H, Chourrout D, Philippe H. Tunicates and not cephalochordates are the closest living relatives of vertebrates. Nature. 2006; 439:965–968. [PubMed: 16495997]

DeVore NM, Scott EE. Structures of cytochrome P450 17A1 with prostate cancer drugs abiraterone and TOK-001. Nature. 2012; 482:116–119. [PubMed: 22266943]

Eramian D, Shen MY, Devos D, Melo F, Sali A, Marti-Renom MA. A composite score for predicting errors in protein structure models. Protein Sci. 2006; 15:1653–1666. [PubMed: 16751606]

Ghosh D, Griswold J, Erman M, Pangborn W. Structural basis for androgen specificity and oestrogen synthesis in human aromatase. Nature. 2009; 457:219–223. [PubMed: 19129847]

Ghosh D, Griswold J, Erman M, Pangborn W. X-ray structure of human aromatase reveals an androgen-specific active site. J Steroid Biochem Mol Biol. 2010; 118:197–202. [PubMed: 19808095]

Gilep AA, Sushko TA, Usanov SA. At the crossroads of steroid hormone biosynthesis: the role, substrate specificity and evolutionary development of CYP17. Biochim Biophys Acta. 2011; 1814:200–209. [PubMed: 20619364]

Goldstone JV, Hamdoun A, Cole BJ, Howard-Ashby M, Nebert DW, Scally M, Dean M, Epel D, Hahn ME, Stegeman JJ. The chemical defensome: environmental sensing and response genes in the *Strongylocentrotus purpuratus* genome. Dev Biol. 2006; 300:366–384. [PubMed: 17097629]

Goldstone JV, McArthur AG, Kubota A, Zanette J, Parente T, Jonsson ME, Nelson DR, Stegeman JJ. Identification and developmental expression of the full complement of Cytochrome P450 genes in Zebrafish. BMC Genomics. 2010; 11:643. [PubMed: 21087487]

Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser. 1999; 41:95–98.

Hines GA, Watts SA, Walker CW, Voogt PA. Androgen metabolism in somatic and germinal tissues of the sea star Asterias vulgaris. Comp Biochem Physiol B. 1992; 102:521–526. [PubMed: 1499289]

Hong Y, Li H, Yuan YC, Chen S. Molecular characterization of aromatase. Ann N Y Acad Sci. 2009; 1155:112–120. [PubMed: 19250198]

Hong Y, Li H, Yuan YC, Chen S. Sequence-function correlation of aromatase and its interaction with reductase. J Steroid Biochem Mol Biol. 2010; 118:203–206. [PubMed: 19944754]

Howard-Ashby M, Materna SC, Brown CT, Chen L, Cameron RA, Davidson EH. Gene families encoding transcription factors expressed in early development of Strongylocentrotus purpuratus. Dev Biol. 2006; 300:90–107. [PubMed: 17054934]

Keay J, Bridgham JT, Thornton JW. The Octopus vulgaris estrogen receptor is a constitutive transcriptional activator: evolutionary and functional implications. Endocrinology. 2006; 147:3861–3869. [PubMed: 16690796]

Lafont R, Mathieu M. Steroids in aquatic invertebrates. Ecotoxicology. 2007; 16:109–130. [PubMed: 17238002]

Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J Biomol NMR. 1996; 8:477–486. [PubMed: 9008363]

Markov GV, Tavares R, Dauphin-Villemant C, Demeneix BA, Baker ME, Laudet V. Independent elaboration of steroid hormone signaling pathways in metazoans. Proc Natl Acad Sci U S A. 2009; 106:11913–11918. [PubMed: 19571007]

Mast N, Annalora AJ, Lodowski DT, Palczewski K, Stout CD, Pikuleva IA. Structural basis for three-step sequential catalysis by the cholesterol side chain cleavage enzyme CYP11A1. J Biol Chem. 2011; 286:5607–5613. [PubMed: 21159775]

Matsumoto T, Nakamura AM, Mori K, Akiyama I, Hirose H, Takahashi Y. Oyster estrogen receptor: cDNA cloning and immunolocalization. Gen Comp Endocrinol. 2007; 151:195–201. [PubMed: 17324427]

Meyer A, Van de Peer Y. From 2R to 3R: evidence for a fish-specific genome duplication (FSGD). Bioessays. 2005; 27:937–945. [PubMed: 16108068]

Miller TW, Shin I, Kagawa N, Evans DB, Waterman MR, Arteaga CL. Aromatase is phosphorylated in situ at serine-118. J Steroid Biochem Mol Biol. 2008; 112:95–101. [PubMed: 18822378]

Miller WL. Androgen biosynthesis from cholesterol to DHEA. Mol Cell Endocrinol. 2002; 198:7–14. [PubMed: 12573809]

Mizuta T, Kubokawa K. Presence of sex steroids and cytochrome P450 genes in amphioxus. Endocrinology. 2007; 148:3554–3565. [PubMed: 17446181]

Mizuta T, Asahina K, Suzuki M, Kubokawa K. In vitro conversion of sex steroids and expression of sex steroidogenic enzyme genes in amphioxus ovary. J Exp Zool A Ecol Genet Physiol. 2008; 309:83–93. [PubMed: 18196586]

Moretti S, Armougom F, Wallace IM, Higgins DG, Jongeneel CV, Notredame C. The M-Coffee web server: a meta-method for computing multiple sequence alignments by combining alternative alignment methods. Nucleic Acids Res. 2007; 35:W645–648. [PubMed: 17526519]

Morrison AM, Goldstone JV, Lamb DC, Kubota A, Lemaire B, Stegeman JJ. Identification, modeling and ligand affinity of early deuterostome CYP51s, and functional characterization of recombinant zebrafish sterol 14alpha-demethylase. Biochim Biophys Acta. 2014; 1840:1825–1836. [PubMed: 24361620]

Nagasawa K, Treen N, Kondo R, Otoki Y, Itoh N, Rotchell JM, Osada M. Molecular characterization of an estrogen receptor and estrogen-related receptor and their autoregulatory capabilities in two Mytilus species. Gene. 2015; 564:153–159. [PubMed: 25862924]

Nakajin S, Shively JE, Yuan PM, Hall PF. Microsomal cytochrome P-450 from neonatal pig testis: two enzymatic activities (17 alpha-hydroxylase and c17,20-lyase) associated with one protein. Biochemistry. 1981; 20:4037–4042. [PubMed: 6793062]

Nelson DR, Goldstone JV, Stegeman JJ. The cytochrome P450 genesis locus: the origin and evolution of animal cytochrome P450s. Philos Trans R Soc Lond B Biol Sci. 2013; 368:20120474. [PubMed: 23297357]

Pallan SP, Nagy LD, Lei L, Gonzalez E, Kramlinger VM, Azumaya CM, Wawrzak Z, Waterman MR, Guengerich FP, Egli M. Structural and Kinetic Basis of Steroid 17α,20-Lyase Activity in Teleost Fish Cytochrome P45017A and Its Absence in Cytochrome P450 17A2. J Biol Chem. 2015 in press.

Parajes S, Griffin A, Taylor AE, Rose IT, Miguel-Escalada I, Hadzhiev Y, Arlt W, Shackleton C, Muller F, Krone N. Redefining the initiation and maintenance of zebrafish interrenal

steroidogenesis by characterizing the key enzyme cyp11a2. Endocrinology. 2013; 154:2702–2711. [PubMed: 23671259]

Payne AH, Hales DB. Overview of steroidogenic enzymes in the pathway from cholesterol to active steroid hormones. Endocr Rev. 2004; 25:947–970. [PubMed: 15583024]

Petrunak EM, DeVore NM, Porubsky PR, Scott EE. Structures of Human Steroidogenic Cytochrome P450 17A1 with Substrates. J Biol Chem. 2014; 289:32952–32964. [PubMed: 25301938]

Pikuleva IA. Cholesterol-metabolizing cytochromes P450. Drug Metab Dispos. 2006; 34:513–520. [PubMed: 16434543]

Schoenmakers HJ, Voogt PA. In vitro biosynthesis of steroids from androstenedione by the ovaries and pyloric caeca of the starfish Asterias rubens. Gen Comp Endocrinol. 1981; 45:242–248. [PubMed: 6945284]

Schoenmarkers HJ, Voogt PA. In vitro biosynthesis of steroids from progesterone by the ovaries and pyloric ceca of the starfish Asterias rubens. Gen Comp Endocrinol. 1980; 41:408–416. [PubMed: 6967842]

Sethumadhavan K, Bellino FL, Thotakura NR. Estrogen synthetase (aromatase). The cytochrome P-450 component of the human placental enzyme is a glycoprotein. Mol Cell Endocrinol. 1991; 78:25–32. [PubMed: 1936523]

Shen MY, Sali A. Statistical potential for assessment and prediction of protein structures. Protein Sci. 2006; 15:2507–2524. [PubMed: 17075131]

Shimozawa O, Sakaguchi M, Ogawa H, Harada N, Mihara K, Omura T. Core glycosylation of cytochrome P-450(arom). Evidence for localization of N terminus of microsomal cytochrome P-450 in the lumen. J Biol Chem. 1993; 268:21399–21402. [PubMed: 8407981]

Solovyev V, Kosarev P, Seledsov I, Vorobyev D. Automatic annotation of eukaryotic genes, pseudogenes and promoters. Genome Biol. 2006; 7(1):S10 11–12. [PubMed: 16925832]

Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014; 30:1312–1313. [PubMed: 24451623]

Storbeck KH, Swart P, Swart AC. Cytochrome P450 side-chain cleavage: insights gained from homology modeling. Mol Cell Endocrinol. 2007:265–266. 65–70.

Strushkevich N, MacKenzie F, Cherkesova T, Grabovec I, Usanov S, Park HW. Structural basis for pregnenolone biosynthesis by the mitochondrial monooxygenase system. Proc Natl Acad Sci U S A. 2011; 108:10139–10143. [PubMed: 21636783]

Tchoudakova A, Callard GV. Identification of multiple CYP19 genes encoding different cytochrome P450 aromatase isozymes in brain and ovary. Endocrinology. 1998; 139:2179–2189. [PubMed: 9529008]

Thornton JW, Need E, Crews D. Resurrecting the ancestral steroid receptor: ancient origin of estrogen signaling. Science. 2003; 301:1714–1717. [PubMed: 14500980]

Vogeler S, Galloway TS, Lyons BP, Bean TP. The nuclear receptor gene family in the Pacific oyster, Crassostrea gigas, contains a novel subfamily group. BMC Genomics. 2014; 15:369. [PubMed: 24885009]

Voogt PA, den Besten PJ, Kusters GC, Messing MW. Effects of cadmium and zinc on steroid metabolism and steroid level in the sea star Asterias rubens L. Comp Biochem Physiol C. 1987; 86:83–89. [PubMed: 2881732]

Wada A, Waterman MR. Identification by site-directed mutagenesis of two lysine residues in cholesterol side chain cleavage cytochrome P450 that are essential for adrenodoxin binding. J Biol Chem. 1992; 267:22877–22882. [PubMed: 1429635]

Webb B, Sali A. Comparative Protein Structure Modeling Using MODELLER. Curr Protoc Bioinformatics. 2014; 47:561–5632.

Wilson JY, McArthur AG, Stegeman JJ. Characterization of a cetacean aromatase (CYP19) and the phylogeny and functional conservation of vertebrate aromatase. Gen Comp Endocrinol. 2005; 140:74–83. [PubMed: 15596073]

Yeh RF, Lim LP, Burge CB. Computational inference of homologous gene structures in the human genome. Genome Res. 2001; 11:803–816. [PubMed: 11337476]

Zanette J, Goldstone JV, Bainy AC, Stegeman JJ. Identification of CYP genes in *Mytilus* (mussel) and *Crassostrea* (oyster) species: first approach to the full complement of cytochrome P450 genes in bivalves. Mar Environ Res. 2010; 69(1):S1–3. [PubMed: 19926125]

Zhou LY, Wang DS, Kobayashi T, Yano A, Paul-Prasanth B, Suzuki A, Sakai F, Nagahama Y. A novel type of P450c17 lacking the lyase activity is responsible for C21-steroid biosynthesis in the fish ovary and head kidney. Endocrinology. 2007a; 148:4282–4291. [PubMed: 17569754]

Zhou LY, Wang DS, Shibata Y, Paul-Prasanth B, Suzuki A, Nagahama Y. Characterization, expression and transcriptional regulation of P450c17-I and -II in the medaka, Oryzias latipes. Biochem Biophys Res Commun. 2007b; 362:619–625. [PubMed: 17723227]

Abbreviations

Adx	adrenodoxin
CYP	cytochrome P450 monooxygenase
HSD	hydroxysteroid dehydrogenase
PREG	pregnenolone
DHEP	dehydroepiandrostenedione
AND	androstenedione
PRG	progesterone
TEST	testosterone
POR	cytochrome P450 reductase

Highlights

- Vertebrate steroid hormone synthesis pathways are 100s of millions of years old.
- Relevant steroidogenic cytochrome P450s evolved sequentially in the deuterostomes.
- CYP11-like genes preceded the appearance of CYP19 genes.
- Substrate binding residues are highly conserved in each steroidogenic CYP family.

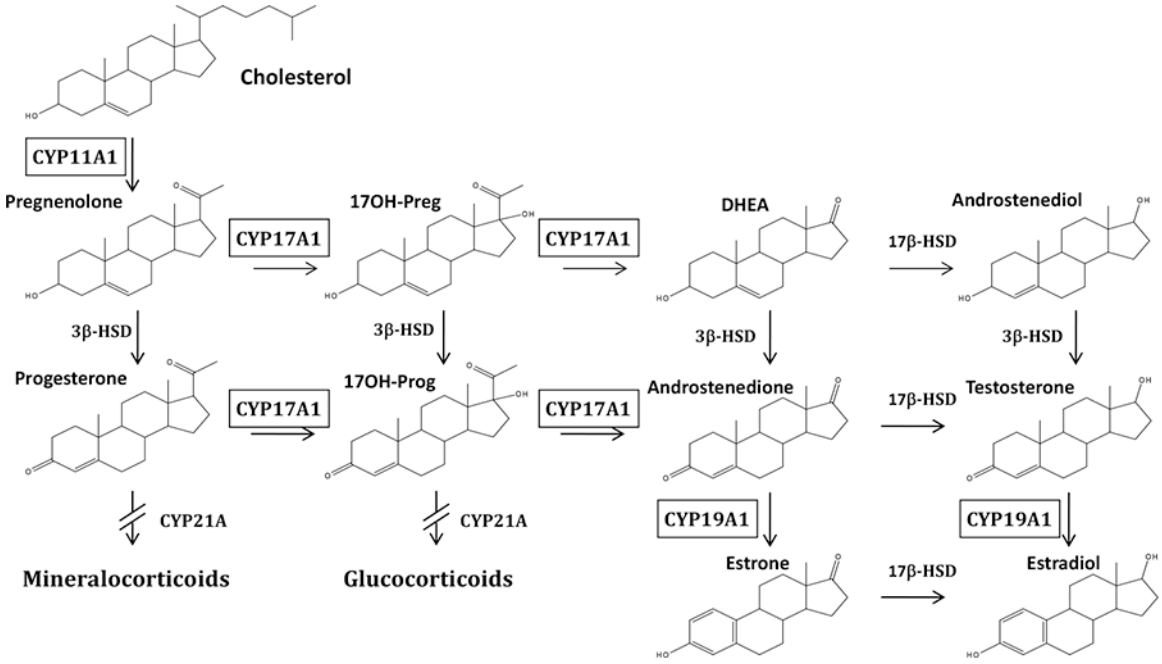
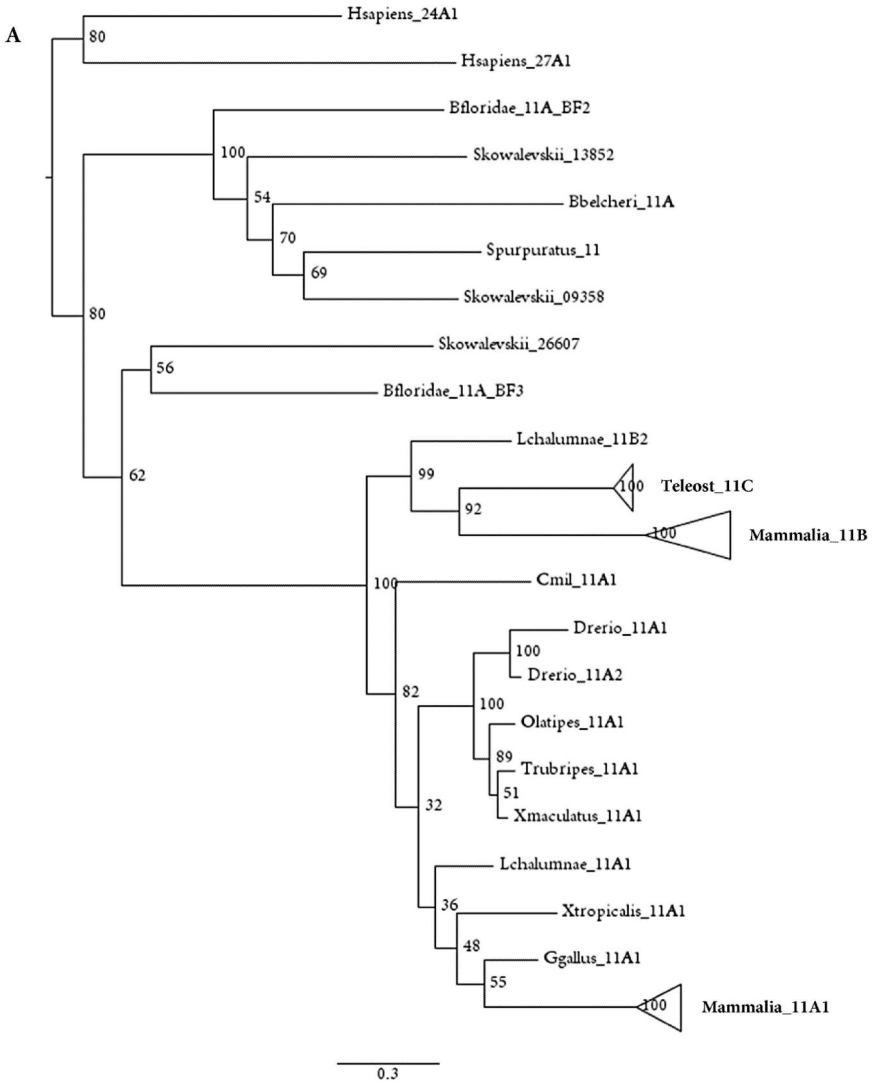
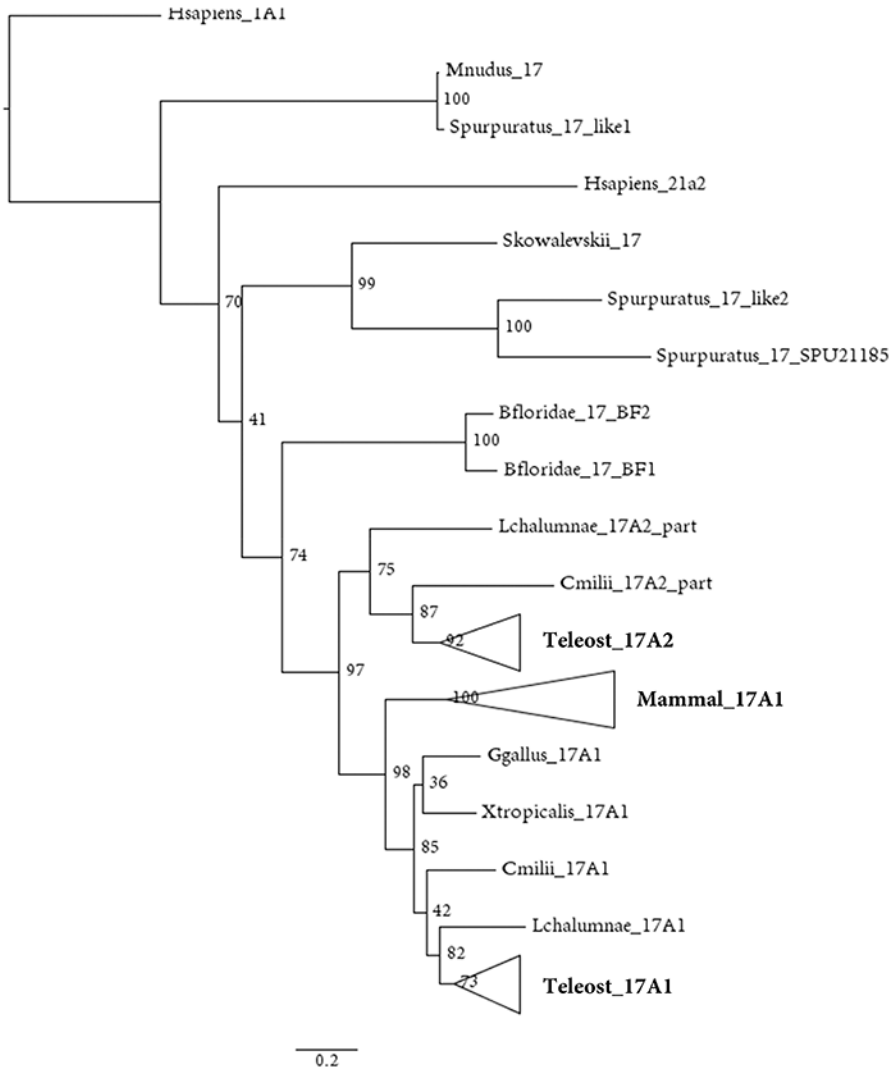


Figure 1.
Sex hormone biosynthesis pathway in vertebrates.



B



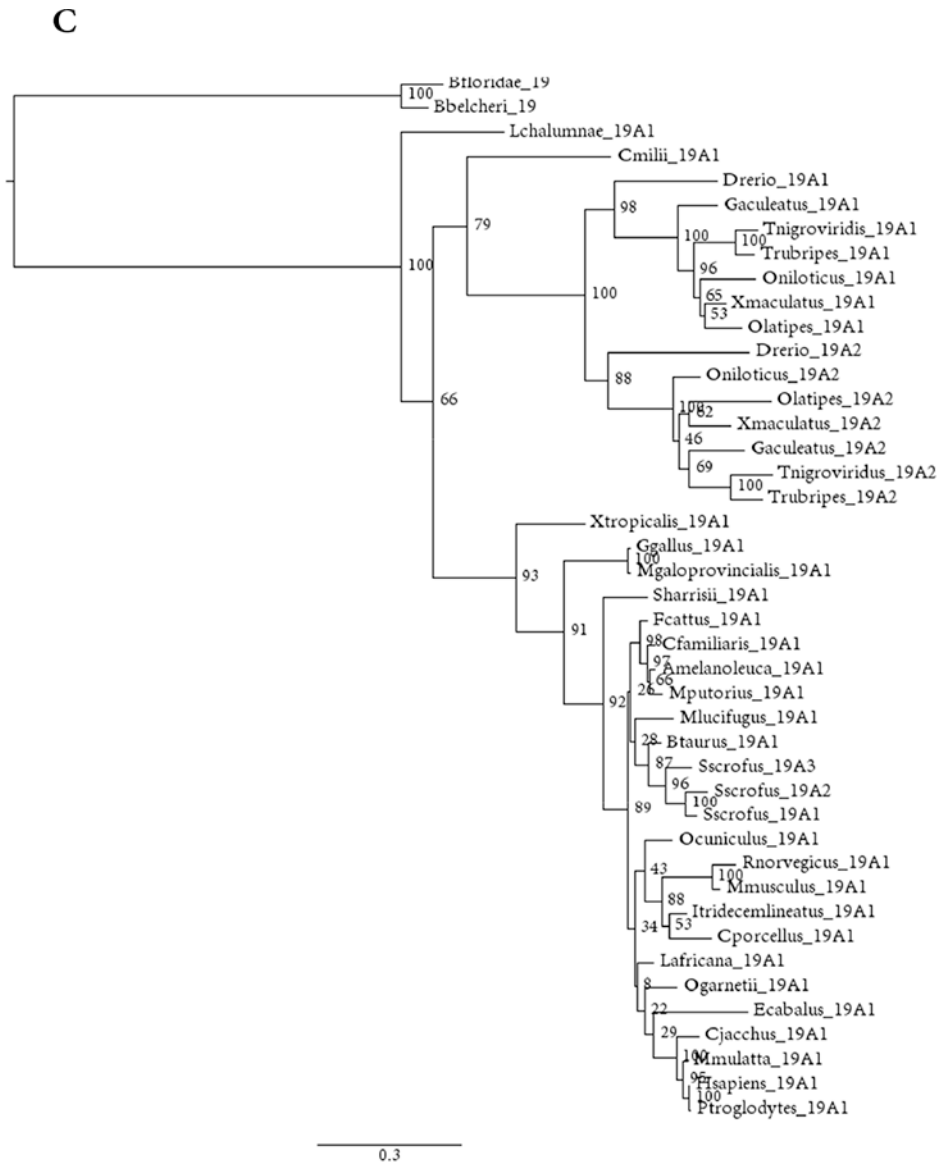


Figure 2.

Maximum likelihood phylogenetic trees constructed from subsets of the largest alignments. A. CYP11A B. CYP17A. C. CYP19A. Note in Figure 2A the invertebrate CYP11 clade diverging prior to the vertebrate CYP11A/B/C clade, with the inclusion of an echinoderm (Spurpuratus_ 11A-like) and two hemichordate CYP11-like (Skowalevski_26607, Skowalevski_09358) sequences with the amphioxus CYP11s (Bfloridae_11BF2, Bbelcheri_11). In Figure 2B, minimal support is found for several sea urchin sequence (Spurpuratus_SPU21185, Spurpuratus_17like2) clustering with the chordate CYP17s. Note that CYP11A activity has been measured in sea urchins, and CYP17A activity measured in sea stars (see text). No non-chordate CYP19 has been found. See Supplemental Tables 1 and 3 for a complete list of species abbreviations and sequence accession numbers.

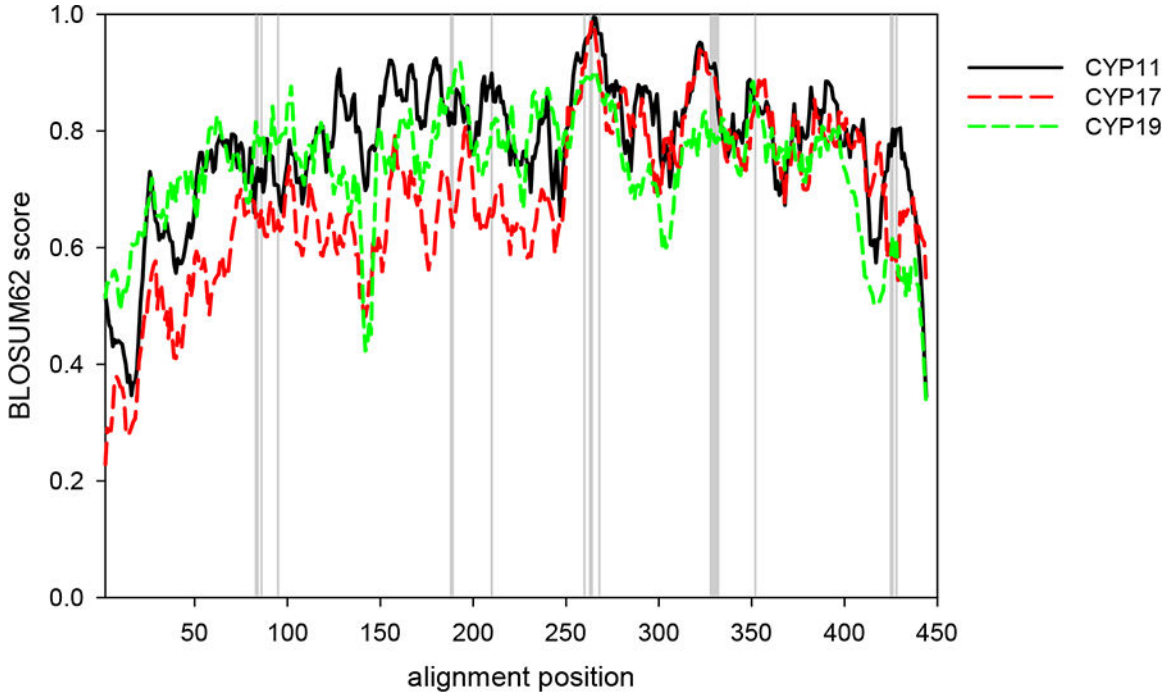


Figure 3.

Sequence similarity across CYP alignments BLOSUM62 sequence similarity across the large CYP alignments, including the additional isoforms (CYP19A2 and CYP17A2) present in fish. The values are seven-residue running averages of the BLOSUM62 score. The bars on the plot indicate the location of the substrate binding residues for CYP11A and CYP17A in the alignment, showing the increased similarity in the regions adjacent to the substrate contact points.

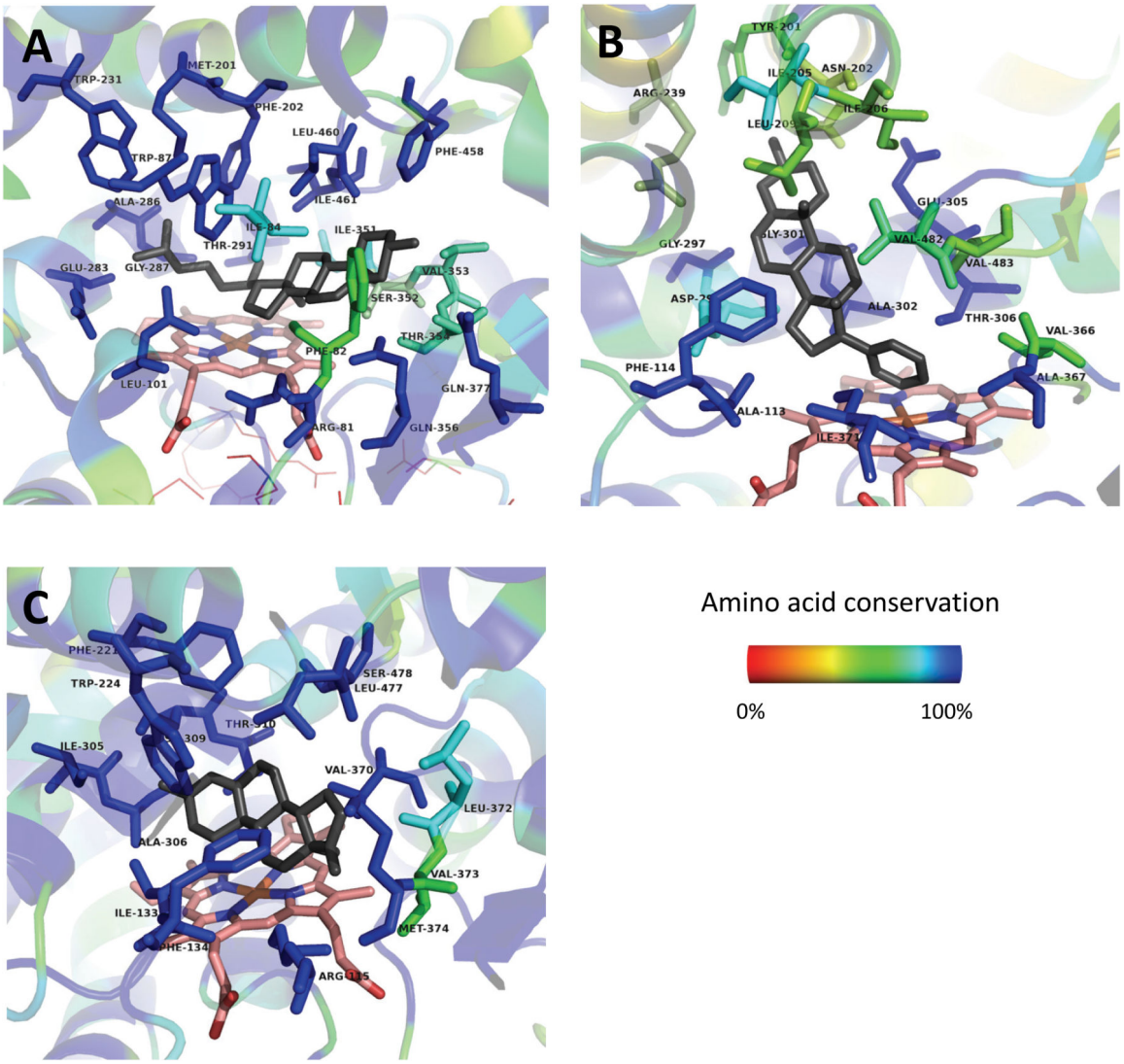


Figure 4.

Substrate binding sites of (A) human CYP11A1, (B) human CYP17A1, and (C) human CYP19A1. The heme group is shown in tan, the substrate molecules are shown grey and the surrounding protein residues are labeled. The amino acids residues within 5A of the bound ligand are shown, colored by conservation score (JTT/PET91 score), with blue representing fully conserved residues. Note that nearly all of the residues within 5 A of the substrate are nearly or completely conserved across the selected deuterostome (CYP11 and CYP17) or chordate (CYP19) sequences.