Evolutionary origins of the estrogen signaling system: insights from amphioxus

Note: on lines 471 and 708 changes made to Genbank # HQ115077 to HQ010363. Be sure to change on final version before publication.

*[a]Callard GV, [c]Tarrant AM, [a]Novillo A, [a]Yacci P, [a]Ciaccia L, [b]Vajda S, [b]Chuang G-Y, Kozakov D, [a]Greytak SR, [a]Sawyer S, [a]Hoover C, and [a]Cotter K

Departments of [a]Biology and [b]Biomedical Engineering

Boston University

5 Cummington St

Boston MA 02215

and

[c]Department of Biology

Woods Hole Oceanographic Institute

Woods Hole MA

***Corresponding author**

Key words: estrogen: aromatase; estrogen receptors; evolution; amphioxus

ABSTRACT

Classically, the estrogen signaling system has two core components: cytochrome P450 aromatase (*CYP19*), the enzyme complex that catalyzes the rate limiting step in estrogen biosynthesis; and estrogen receptors (ERs), ligand activated transcription factors that interact with the regulatory region of target genes to mediate the biological effects of estrogen. While the importance of estrogens for regulation of reproduction, development and physiology has been well-documented in gnathostome vertebrates, the evolutionary origins of estrogen as a hormone are still unclear. As invertebrates within the phylum Chordata, cephalochordates (e.g. the amphioxus of the genus *Branchiostoma*) are among the closest invertebrate relatives of the vertebrates and can provide critical insight into the evolution of vertebrate-specific molecules and pathways. To address this question, this paper briefly reviews relevant earlier studies that help to illuminate the history of the aromatase and ER genes, with a particular emphasis on insights from amphioxus and other invertebrates. We then present new analyses of amphioxus aromatase and ER sequence and function, including an *in silico* model of the amphioxus aromatase protein, and *CYP19* gene analysis. *CYP19* shares a conserved gene structure with vertebrates (9 coding exons) and moderate sequence conservation (40% amino acid identity with human *CYP19*). Modeling of the amphioxus aromatase substrate binding site and simulated docking of androstenedione in comparison to the human aromatase shows that the substrate binding site is conserved and predicts that androstenedione could be a substrate for amphioxus *CYP19*. The amphioxus ER is structurally similar to vertebrate ERs, but differs in sequence and key residues of the ligand binding domain. Consistent with results from other laboratories, amphioxus ER did not bind radiolabeled estradiol, nor did it modulate gene expression on an

estrogen-responsive element (ERE) in the presence of estradiol, 4-hydroxytamoxifen, diethylstilbestrol, bisphenol A or genistein. Interestingly, it has been shown that a related gene, the amphioxus "steroid receptor" (SR), can be activated by estrogens and that amphioxus ER can repress this activation. *CYP19*, *ER* and *SR* are all primarily expressed in gonadal tissue, suggesting an ancient paracrine/autocrinesignaling role, but it is not yet known how their expression is regulated and, if estrogen is actually synthesized in amphioxus, whether it has a role in mediating any biological effects . Functional studies are clearly needed to link emerging bioinformatics and *in vitro* molecular biology results with organismal physiology to develop an understanding of the evolution of estrogen signaling.

1. INTRODUCTION

Based primarily on evidence from humans and laboratory mammals, it is well established that estrogens play a critical regulatory role in many different life processes beginning in early stages of embryogenesis. The term "estrogen" derives from its first perceived function as a female reproductive hormone, specifically associated with the period of sexual receptivity in female mammals (estrus = Latin *oestrus* meaning frenzy or gadfly). Although early investigators used the urine of pregnant women to isolate estrone, the first steroid found to have hormonal activity, subsequent studies soon reported the presence of estrogens and the biosynthesis of estradiol, estrone and estriol from small acyclic precursors in both males and females of a wide range of vertebrates from fish to mammals [1]. It is now generally accepted that estrogen not only is required for the normal growth, development and functioning of the reproductive system but also has a critical role in diverse other tissue types and organ systems, including brain, bone, skin, fat, cardiovascular and metabolic.. Excesses or deficiencies of estrogen are associated with various pathological states, such as breast and prostate cancer and osteoporosis. Environmental chemicals that are estrogen-like in their bioactivity have been implicated in developmental abnormalities and endocrine-disrupting effects in humans and animals. Not surprisingly, factors and mechanisms regulating estrogen production and signal transduction continue to be a matter of intense research interest (reviewed by [2,3]).

Classically, the estrogen signaling system has two core components: cytochrome P450 aromatase, the enzyme complex that catalyzes the rate limiting step in estrogen biosynthesis; and estrogen receptors (ERs), ligand activated transcription factors that interact with the regulatory region of target genes to mediate the biological effects of estrogen. While this viewpoint continues to serve as a valuable template for basic and clinical studies, advances in molecular

93 endocrinology reveal that the complexity and diversity of estrogen physiology is accomplished
94 by multiple signaling modes (endocrine, paracrine, autocrine/intracrine), as defined by the
95 nature, proximity and topographical relationship of aromatase and ER expressing cells; two or
96 more genetically distinct ER subtypes and multiple ER splice variants; diverse other classes of
97 membrane- and nuclear-localized receptors; and an array of different cellular signal transduction
98 pathways (genomic, nuclear-mediated; non-genomic/membrane-mediated)(see section 1.2.1,
99 below).

100 Fundamental questions remain regarding the evolution of the estrogen mediated signaling
101 system. What are the evolutionary origins and molecular nature of the core components
102 (aromatase and ER)? Which receptor signal transduction pathway is most ancient? Is the original
103 messenger molecule the endogenously synthesized estrogen we know in vertebrates (estradiol,
104 estrone)? Or did estrogen-like environmental molecules have the earliest signaling role? The
105 basic anatomy, physiology and biochemistry of estrogen signaling have been extensively studied
106 in representatives of all major groups of jawed vertebrates, signifying an ancient and
107 evolutionarily conserved regulatory role. More recently, the structures and phylogenetic
108 distribution of genes encoding aromatase (Figure 1a, [4,5]) and ER (Figure 1b, [6-10]) have
109 been documented, reinforcing the earlier work, but mechanistic details of estrogen-mediated
110 signaling in organisms that predate the gnathostomes is not entirely clear. One approach to
111 addressing the question is to study the closest invertebrate relatives of vertebrates and to
112 determine precursors of vertebrate-specific molecules and pathways in these organisms. In
113 addition to vertebrates, the phylum Chordata includes two invertebrate groups: urochordates (e.g.
114 the ascidian *Ciona intestinalis*) and cephalochordates (e.g. the amphioxus of the genus
115 *Branchiostoma*). In this paper, we briefly review the evolutionary history of the aromatase and

ER genes, with a particular emphasis on insights from amphioxus and other invertebrates, and then present new analyses of aromatase and ER in amphioxus.

1.1 Cytochrome P450 aromatase and the *CYP19* gene

1.1.1. Structure and function

The critical enzyme for estrogen synthesis is aromatase, a member of the cytochrome P450 (*CYP*) superfamily of monooxygenase enzymes [11]. The membrane-associated aromatase complex catalyzes the transformation of androgens (androstenedione and testosterone) to estrogens (estradiol and estrone) and is the product of a single *CYP19A1* gene in humans. Although most highly expressed in estrogen secreting glandular tissues, such as placenta and gonads, aromatase is expressed in a wide array of other tissue types: brain, fat, bone, pituitary in humans; brain, pituitary, retina in teleost fish. Of these, certain cell/tissue types are competent to transform acyclic precursors stepwise through cholesterol all the way to estrogen (ovary), whereas others are competent in the final aromatization step but are lacking one or more of the earlier enzymes in the steroidogenic pathway. Human placenta, for example, lacks C17,20 lyase (*CYP17*) and relies on androgen precursors supplied by the fetal adrenal for estrogen production.

The aromatase protein is monomeric and is anchored within the endoplasmic reticulum by a membrane-spanning region of the amino terminus [12,13]. The crystal structure of the human aromatase protein has recently been determined [14]. The 503-residue polypeptide chain folds into 12 major α-helices and 10 β-strands and forms a heme group and adjacent steroid binding site near the geometric center of the protein [15]. This overall folding pattern is similar to other membrane-bound P450s, and several regions show strong sequence conservation including helices H-K, the aromatic region and especially the heme-binding region. Of the

conserved helices, the "I-helix" is particularly important because it contains several hydrophobic residues that help to form the catalytic cleft and incorporates a key bend at Pro308 that provides additional space to accommodate a steroid substrate [15,16].

1.1.2. Phylogenetic context

Aromatase activity (for review [17]) and the *CYP19* gene(s) have been well-documented in all major classes of gnathostome (jawed) vertebrates. The *CYP19* gene has undergone independent duplications in several lineages, most notably the teleost fish [18,19] and suiform mammals [20,21]. Whereas the teleostean gene duplicates are thought to reflect a whole genome duplication event [22], the three *CYP19* genes of pigs are the result of much more recent tandem duplication events. Duplicate aromatases retain the ability to synthesize estrogens but also exhibit functional differences. Within the teleost fish, duplicated *CYP19* genes differ dramatically in their tissue expression patterns [19,23] as well as in their relative affinity for different androgen and inhibitor substrates [24,25] and inducibility by estrogens and xenoestrogens [18,23,26,27]. Similarly, in suiform mammals, duplicated aromatase genes differ in expression patterns, substrate affinity and product formation [20,21]. While humans possess only a single *CYP19* gene, expression is regulated by 11 promoters and alternative first exons, which are used in a tissue specific manner [28,29]. Along with the diverse roles played by estrogens, this complexity of aromatase regulation indicates the importance and richness of the estrogen signaling pathway.

Phylogenetic analyses of the *CYP* superfamily have not revealed close relationships of *CYP19* with any other family members [4,30]; thus, it is not currently possible to trace the origin of aromatase activity from ancestral *CYPs* that served other metabolic functions. *CYP19* orthologs

have recently been identified within amphioxus [4,5]. However, *CYP19* has not been identified within the sequenced genomes of urochordates, echinoderms, or protostomes, nor have they been identified outside of the bilaterian animals [31,32]. Although we cannot rule out the possibility that a recognizable ancestral *CYP19-like gene* or *CYP19* itself was secondarily lost in these groups, the cephalochordate lineage represents the earliest known occurrence of *CYP19* to date. In addition to *CYP19,* amphioxus contains orthologs of other enzymes in the steroidogenic sequence leading to estrogen biosynthesis: *CYP17,* and 17β-hydroxysteroid dehydrogenase [5,33]. In addition, Amphioxus contains *CYP11*-like genes that, along with some uncharacterized cnidarian and placozoan *CYPs,* are positioned as an outgroup to the vertebrate *CYP11* clade [5,31]. *CYP11A* catalyzes cleavage of the side chain from the sterol D-ring; side chain cleavage by *CYP11A* (or a functional equivalent) is necessary for *de novo* synthesis of steroids. Because the catalytic activities of the amphioxus *CYP11*-like genes have not been determined and side-chain cleavage has not been documented, it remains unclear whether amphioxus can synthesize steroids from sterol precursors.

Measurements of steroidogenic activity using radiolabeled precursors and steroid-like immunoreactivity in amphioxus are consistent with the molecular studies described above. Aromatase activity in amphioxus was first demonstrated through the conversion of tritiated 19-hydroxyandrostenedione to estrone and estradiol by homogenates of body segments containing gonads [34]. Interestingly, activity was not detected in homogenates of brain or tail segments. Mizuta and colleagues [35] similarly measured estrogen synthesis by amphioxus ovarian homogenates and documented a suite of steroidogenic conversions. Estrogen synthesis primarily occurred in mature ovarian tissues prior to spawning. Estradiol-like, as well as progesterone- and testosterone-like molecules, have been quantified in amphioxus gonads using radioimmunoassay

[5]. Similar to the patterns in aromatase activity, immunoactive estrogen was present in both ovaries and testes, but not in non-gonadal extracts, and concentrations in the ovary were greatest prior to spawning [5].

1.2. ERs and *Esr* genes

1.2.1. Structure and function

In vertebrates, the classical mechanism of estrogen signaling occurs through specific binding of estradiol to ERs , which are are encoded by *Esr* genes. Within the nuclear receptor superfamily, the ERs form a family with two other receptor groups: the estrogen-related receptors (ERRs), and other vertebrate-type steroid receptors (SRs, which include androgen receptors, progesterone receptors, and corticoid receptors). The human genome contains two ERs, ERα (NR3A1, *Esr1* [36]) and ERβ (NR3A2, *Esr2* [37]), due to a duplication of the *Esr* gene early in the vertebrate lineage [38]. Unique among the vertebrates, however, teleost fish have one ERα but two ERβs (ERβa and ERβb) .

Like other nuclear receptors, ERs have a modular structure divided into key functional domains (A-F) [39]. At the amino terminus, the A/B domains contain the ligand-independent AF-1 activation function [40]. The DNA-binding domain (DBD, C domain) is the most highly conserved region and contains two zinc fingers that enable binding of the ER to specific estrogen responsive elements (EREs) on the DNA. The hinge region (D-domain) has a more variable sequence, contains a nuclear localization signal, and enables synergism between the activation functions (AF-1 and AF-2) for full transcriptional activity [41]. At the amino terminus, the ligand binding domain (E/F) LBD is highly conserved, and serves to bind ligands, enable dimerization, recruit co-factors and stimulate transcription through the ligand-dependent AF-2 region.

In the absence of ligand, ERs generally occur in complexes with chaperones, such as Hsp90 [42]. Upon binding of estradiol or another agonist, ERs dissociate from the chaperones, form homo- or heterodimers [43]), recruit cofactors, bind to DNA and modulate transcription of target genes. Utilization of multiple promoters and alternative splicing creates additional complexity in ER signaling. Eight promoters have been identified for human ERα and two for ERβ, which function in tissue-specific expression [44-47]. Alternate splicing generates an exceptional number of ER isoforms lacking one or more functionally important domains; these variants differ in their expression patterns and functional properties [47]. For example, a human ERβ isoform (ERβcx) truncated at the C-terminus has been reported heterodimerize with wild-type ERβ and function as a dominant negative [47-49].

In addition to modulating the activity of nuclear receptors, steroids can also stimulate rapid cellular responses which are mediated through membrane-bound receptors [50,51]. With respect to estrogen signaling, rapid effects have been attributed to interactions with classical nuclear ERs that are localized within the cell membrane [52-54] as well as with GPR30, a G-protein coupled receptor [55]. To date, membrane-bound ERs have only been rigorously characterized in mammals and fish [56,57]. Estrogens have been shown to exert similar rapid effects on cell signaling in molluscs [58]; however, the genes encoding membrane-bound ERs have not yet been identified in invertebrates, and it has not yet been demonstrated that estradiol is the endogenous activator of this receptor.

1.2.2. Phylogenetic context

ERs have been identified and shown to be activated by steroidal estrogens in all classes of vertebrates, including the agnathan sea lamprey [6]. Among invertebrates, homologs to the

ERs have been identified in amphioxus [7,33] as well as in molluscs [9,59] and annelids [10]. Previous phylogenetic analyses conducted using a variety of methods (parsimony, likelihood, Bayesian) have shown that chordate ERs (vertebrate and amphioxus) form a clade [7,10] and that the protostome ERs (mollusc and annelid) comprise a sister group [9,10]. In addition, Keay and Thornton [10] found that this bilaterian ER clade was supported as a sister group to the SRs. In their study, the position of the protostomes ERs was only moderately supported, but much of the observed uncertainty could be attributed to the effects of a long branch associated with the amphioxus SR.

As demonstrated by reporter assays in mammalian cell lines, ERs from amphioxus [6,8,60] and from molluscs [9,59] are not activated by steroidal estrogens. In contrast, ERs from two annelid species bind estrogens with high affinity and activate transcription in response to low doses (EC50 < 10 nM estradiol) of estrogens [10], although it remains to be determined whether steroidal estrogens are physiological ligands for these annelid receptors. Based on phylogenetic patterns and reconstructions of predicted ancestral receptors, it has been hypothesized that the ancestral ER originated early in the bilaterian lineage and was activated by estrogens ([10,61], but see also [6,31]). One interpretation is that ER activation by estrogens was a property that was lost within the lineage leading to the cephalochordates and that the ER gene per se was lost from echinoderms, urochordates and several protostome lineages.

Within the large nuclear receptor superfamily (48 genes in human, 33 in amphioxus [33]), the ERs form a family (NR3A) with two other receptor groups: the estrogen-related receptors (ERRs, NR3B), and other steroid receptors (SRs, NR3C, which include androgen receptors, progesterone receptors, and corticoid receptors). Amphioxus has one representative gene in each of these three groups [7,33]. As mentioned above, cell-based reporter assays

indicate the amphioxus ER ortholog does not stimulate transcription of ERE-driven reporters or interact with the coactivator SRC-1 in response to estradiol. Somewhat surprisingly (but as hypothesized by Paris and colleagues [6]), reporter assays indicate that the amphioxus SR stimulates transcription through EREs and AREs (androgen-responsive elements) in response to estradiol and estrone [8,60]. Amphioxus ER and SR share overlapping affinities for DNA binding sites, and reporter assays indicate that ER can competitively repress estradiol-induced signaling by SR [8] as well as by human ERα and ERβ [6]. Binding of ligands to amphioxus ER was not directly measured in these studies, but limited proteolysis assays suggested that the amphioxus ER is unlikely to bind estradiol or several other ligands for vertebrate ERs [6]. Cell-based reporter assays have been used to screen a variety of ligands (e.g., 3β-androstenediol, resveratrol, enterolactone, diethylstilbestrol [6]) for their ability to modulate signaling by amphioxus ER, but no functional ligands have been identified. Interestingly, although limited proteolysis assays suggested that the plasticizer bisphenol A can bind amphioxus ER, this ligand did not affect transactivation [6].

Bridgham and colleagues [8] noted that 11 of the 18 residues that line the ligand-binding pocket of human ERα are altered in amphioxus ER, but only 4 of 18 in amphioxus SR. Through comparison with the human ERα crystal structure, they identified two key substitutions likely to disrupt hydrogen bonding and packing interactions that would normally stabilize the ligand within the binding pocket in a trancriptionally active conformation. They then conducted site-directed mutagenesis, and experimentally demonstrated that the two substitutions (corresponding to amino acids 394 and 404 in the LBD of human ERα) are indeed sufficient to confer repressive activity on the SR.

As part of a long term program of research in this laboratory that focuses on the origin and evolution of estrogen signaling in vertebrates, we sought to obtain insights by studying aromatase and ER in amphioxus. Here we confirm and extend studies cited above, and present new information on *CYP19* gene organization, including an *in silico* model of the aromatase protein.

2. MATERIALS AND METHODS

2.1. Animals, treatments, and nucleic acid extraction

Amphioxus (*Branchiostoma floridae*) were purchased from Gulf Specimen Marine Lab (Panacea, FL). Animals were obtained in May, when adults were reproductively active and readily sexed by visualizing the gonads through the transparent body wall. Immediately upon receipt, animals were chilled to 4^{o}C on ice, sexed, and divided into cephalic (anterior to the gonads), caudal (posterior to the gonads), and central (gonad-containing) regions under a dissecting microscope as previously described [34].

Tissues were used to prepare RNA (as in [18,62]) for cloning and semi-quantitative PCR analysis. For analysis of genomic sequence, DNA was extracted from tail segments of individual amphioxus. Briefly tissue (250 mg) was incubated overnight at 56^{o}C in 500 µl of lysis buffer (50 mM Tris-HCl [pH 8.0], 5 mM EDTA [pH 8.0], 200 mM NaCl, 1% [w/v] sodium dodecyl sulfate containing proteinase K to a final concentration of 0.1 mg/mL). After addition of 500µl isopropanol, the sample was centrifuged for 5 min at 3500 rpm at 4 ^{o}C. The resulting DNA pellet was washed once with 100% ethanol (1 ml) and once with 75% ethanol (1 ml), air dried for 10 min, and resuspended in 30 µl TE buffer (10 mM Tris- HCl/1 mM EDTA).

2.1 cDNA cloning and analysis

Using total RNA from ovarian segments and methods previously described in detail for teleostean cDNAs [63,64] , amphioxus aromatase and ER cDNAs were amplified stepwise by RT-PCR and 5'- and 3'-RACE. Oligonucleotide primers are shown in Supplementary Table 1. In the case of aromatase, initial primers were designed to target sequences in an *in silico* P450 aromatase predicted by Nelson [65]. For cloning of ER, initial primer sequences were designed to amplify a portion of the ER detected bioinformatics queries of the amphioxus whole genome database using the discontinuous megaBLAST algorithm with human ERα (NM_000125) and ERβ (X99101), Aplysia ER (AY327135) and lamprey ER (AY028456) The sequence identified as a putative amphioxus DBD was extended in the 3' and 5' directions using an *in silico* DNA-walking approach in combination with 5' and 3'-RACE.

For both aromatase and ER, full coding sequences were then amplified as single products, confirming assembly of the cDNA fragments. Deduced aromatase and ER sequences were aligned using Clustal W with sequences previously reported from representative vertebrate taxa (Accession numbers shown in Fig 1 caption). To confirm the phylogenetic relationship of the cloned amphioxus sequences, trees were constructed using Neighbor-Joining and/or maximum likelihood criteria. For *Cyp19*, the tree was rooted using the human *Cyp17* and *Cyp21* sequences, which are both members of the *Cyp2* clan [30,65]. A maximum likelihood tree was constructed using RAXML [66] with a WAG matrix (selected by AIC using ProtTest version 2.4 [67]) and 100 bootstrap replicates. For ER, a Neighbor-Joining tree was constructed in Phylip 3.6 [68] with 1000 bootstrap replicates and a PAM Dayhoff matrix.

2.2. Genomic DNA cloning and sequence analysis

Intronic sequence was obtained for the *CYP19* gene by PCR amplification of genomic DNA using primers which were specific for sequences in adjacent exons or spanning exon-intron junctions (Supplementary Table 1). 5'-flanking sequence was amplified from genomic DNA using a forward primer targeting genomic sequence and a reverse primer targeting a sequence downstream of putative translational start site in the second exon. Putative cis regulatory elements were identified within the 5'-flanking sequence by comparison with the TRANSFAC database using MATCH with default parameters [69].

2.3. Molecular analysis of amphioxus aromatase

2.3.1. Modeling

The crystal structure of the human aromatase protein has recently been determined [14] and is available in the Protein Data Bank [70], PDB code 3EQM. We used the homologous extension program MODELLER [71,72] to generate a model of amphioxus aromatase. After specifying the target sequence (GenBank ID DQ165086.1), the template sequence and structure (PDB code 3EQM), and an alignment of the two sequences, MODELLER was used to automatically build a 3-dimensional protein model containing all non-hydrogen atoms. The model was refined using energy minimization within MODELLER.

2.3.2. Mapping of aromatase structures

The main goal of constructing a model of the amphioxus aromatase was to compare the binding sites of the human and amphioxus proteins. The comparison uses a very sensitive tool called computational solvent mapping [73,74], originally developed for the identification of "hot spots", i.e., pockets of a protein that bind a variety of small organic molecules. An established

experimental approach to finding such hot spots is screening for the binding of fragment-sized organic compounds [75,76]. Since the binding is very weak, it is usually detected by nuclear magnetic resonance (SAR by NMR [75]) or by X-ray crystallography [76] methods. The FTMAP solvent mapping algorithm used here is a computational analog of the screening experiments, and has been described previously [74]. FTMAP places molecular probes, small organic molecules containing various functional groups, around the protein surface on a dense grid, finds favorable positions by further search using empirical free energy functions, clusters the low energy conformations, and ranks the clusters on the basis of the average free energy. We used 16 small molecules as probes (ethanol, isopropanol, tert-butanol, acetone, acetaldehyde, dimethyl ether, cyclohexane, ethane, acetonitrile, urea, methylamine, phenol, benzaldehyde, benzene, acetamide, and N,N dimethylformamide). The low energy clusters of different probes are further clustered to identify consensus sites, and the importance of such sites is measured in terms of the probe clusters contained. The sites with the largest number of probe clusters are considered as predictions of binding hot spots. Applications to a variety of proteins show that the probes always cluster in important subsites of the binding site and the amino acid residues that interact with many probes also bind the specific ligands of the protein. Since the differences in the number of probe clusters that bind to a particular site highlight even very small conformational changes if those affect the size or surface properties of the pocket, mapping is very useful for comparing homologous proteins or different structures of a protein [77-80]. The comparison is based on residue contact fingerprints. To obtain such fingerprints, the non-bonded interactions and hydrogen bonds between all atoms of the computational probes and the individual protein residues are counted using the HBPLUS program [81].

2.3.3.Docking

After the identification of the important residues in the binding site, we docked androstenedione to both the human aromatase structure and the homology model of the amphioxus aromatase using version 4.0 of the AutoDock program [82]. AutoDock is a suite of automated docking tools. It is designed to predict how small molecules, such as substrates or drug candidates, bind to a receptor of known 3D structure. The docking is restricted to a 40 Å x 40 Å x 40 Å box, centered at the center of the protein. The box is large enough to enclose the entire ligand binding site. Other parameters are assigned the default values given by the AutoDock program. The protein structure is kept fixed during docking. AutoDock employs a genetic algorithm (GA) for conformational sampling, each GA run resulting in a single docked conformation. We performed 100 individual GA runs, thus generating 100 docked conformations for each complex.

2.4. ER binding and transcriptional analysis

The full length Amphioxus ER was subcloned into a v5-tagged expression vector (pcDNA3.1/nV5-DEST, Invitrogen). A similar expression vector was obtained for the human ERα (pcDNA3.1nv5-hERalpha, [83]). To assess the ability of amphioxus ER to bind estradiol, amphioxus ER and human ERα proteins were synthesized using the TnT Quick Coupled Reticulocyte Lysate System (Promega). The specific binding of tritiated estradiol ([6,7-^{3}H] estradiol, 45.0 Ci/mmol, Amersham Biosciences) to *in vitro* expressed ERs was measured using charcoal-based binding assays [84,85]. Briefly, *in vitro* synthesized proteins were diluted in MEEDGM buffer (25 mM MOPS, 1 mM EDTA, 5 mM EGTA, 0.02% NaN_3, 20 mM Na_2MoO_4, 10% (v:v) glycerol, 1 mM DTT, pH 7.5) containing a mixture of protease inhibitors [85]. To correct for variation in expression efficiency, amphioxus ER was diluted 1:10 and human ERα

was diluted 1:20. Aliquots (100 μl) of the diluted proteins were incubated overnight at 4°C with tritiated estradiol in 2.5 μl DMSO. The activity of tritiated estradiol was directly measured in 10 μl from each tube. At the end of the incubation, 30 μl was transferred from each tube in duplicate aliquots to 1.5 ml polypropylene microcentrifuge tubes containing 30 μl of 4 mg/ml dextran-coated charcoal in MEEDGM. Tubes were incubated on ice for 10 min with periodic vortex mixing. The tubes were centrifuged for 2 min at 2000 x g, and activity was quantified in 40 μl of the supernatant by liquid scintillation counting. Nonspecific binding was directly measured using TnT lysate incubated with an empty expression vector [85]. Specific binding of tritiated estradiol to the ERs was calculated by subtracting non-specific binding from total binding. Binding curves were fitted using a one-site binding equation with PRISM software (GraphPad).

Transactivation by amphioxus ER was assessed using a cell-based reporter assay with methods similar to those described by Karchner et al. [86]. COS-7 cells (ATCC) were plated (3 x 10^4 cells/well) in triplicate wells of 48-well plates in phenol red-free MEM (Invitrogen), supplemented with non-essential amino acids, 1 mM sodium pyruvate, 2 mM L-glutamine and 10% charcoal-stripped fetal bovine serum. After 24 hours, cells were transiently transfected using 1 μl Lipofectamine 2000 (Invitrogen) in fresh media along with expression plasmids for an ER (human or amphioxus, 100 ng), a luciferase reporter (3xERE-TATA-LUC, Addgene plasmid 11354 [87], 100 ng) and transfection control (pRL-TK, Promega, 3 ng). The total amount of DNA per well was adjusted to 300 ng through addition of an empty expression vector (pcDNA3.1). Five hours after transfection, cells were treated with vehicle control (0.5% DMSO final concentration), estradiol (1-100 nM), or other potential ligands. Twenty-four hours after transfection, the cells were lysed with passive lysis buffer (Promega), and luminescence was measured using the Dual Luciferase Assay kit (Promega) in a TD 20/20 luminometer (Turner

Designs, Sunnyvale, CA). Transactivation in the presence of DMSO and estradiol was measured in three independent experiments. The other compounds were tested in two independent experiments.

2.6. RT-PCR analysis of aromatase and ER mRNAs

Semi-quantitative RT-PCR was performed using cDNAs from head, gonadal and tail segments from individual amphioxus. Primer sequences are given in Supplementary Table 1. The PCR reactions utilized Platinum Taq polymerase (Invitrogen) according to the manufacturer's instructions. PCR conditions were set to approximate the linear range by optimizing the quantity of input template and cycle number. PCR conditions for aromatase were 94° C/ 5 min, 30 cycles of (94° C/ 30 s, 50°C/ 45 s, and 72°C/ 2 min), followed by 72°C / 10 min. PCR conditions for ER were 94°C for 5 min, 5 cycles of (94°C/ 30 s, 43°C/ 45 s, and 72°C/ 90), then 20 cycles of (94°C/ 30 s, 50°C/ 45 s, and 72°C/ 2 min), followed by 72°C/ 10 min.

3. RESULTS AND DISCUSSION

3.1 Isolation of aromatase cDNA and sequence analysis

The assembled amphioxus *CYP19* cDNA consensus sequence (GenBank Accession number DQ165086) consisted of a single translation initiation site, a 1581 bp open reading frame (ORF) that encoded a predicted protein sequence of 527 aa, and 5′ and 3' UTR of 5 and 1194 bp, respectively. The 3'-UTR terminated in a polyA tail. Compared with the *in silico* sequence initially reported by Nelson [65], our cloned sequence showed 13 overall residue substitutions and a 5 amino acid insertion at the boundary of exons 4 and 5 (amino acid 173, not shown). Two

of the differences were within the conserved I-helix domain. Compared with the partial cDNA sequence reported by Castro and colleagues [4], our sequence contained 3 residue substitutions and a single amino acid insertion (amino acid 373). Our sequence was 88% identical to the *B. belcheri* sequence (433/492 residues). The amino terminus of the *B. floridae CYP19* aromatase is elongated relative to the human and killifish aromatase B sequences and is similar in length to the dogfish and killifish aromatase A sequences. While the *B. belcheri* sequence is not elongated, the predicted start codon aligns with the second methionine in our *B. floridae* sequence. Because no 5'-UTR sequence has been reported for *B. belcheri*, we consider it likely that a portion of the amino terminus has been truncated.

Phylogenetic analysis confirmed that the amphioxus sequence identified in this study is orthologous to the vertebrate aromatases (Fig 1A), consistent with previously published analyses of amphioxus aromatase conducted using neighbor-joining [4,5] and maximum likelihood methods [31]. The tree topology corresponded with the evolutionary relationship between amphioxus and vertebrates [5].

3.2. Isolation of ER cDNA and sequence analysis

The assembled cloned amphioxus ER cDNA (GenBank accession number EF554313.1) contained an ORF of 1383 bp, a 5'-UTR of 684 bp, and two 3'-UTR sequences (988 bp and 633 bp). The long and short UTRs overlapped and were essentially identical in sequence at their 5' ends. Both had polyA tails suggesting they are products of a single mRNA with alternate polyA addition sites. The ORF of the assembled mRNA encoded a polypeptide of 460 aa, and was amplified, cloned and sequenced. The cloned cDNA had >99% identity when compared to the *in silico* derived ER cDNA; however, the protein predicted from the genomic sequence

(JGI_210589), is missing the entire A/B domain 5' of residue 83 of our cloned sequence and contains several indels due to incorrectly predicted exon boundaries. Our cloned sequenced differed by two amino acids from the sequence reported and characterized by Paris et al. [6]: one in the A/B domain (histidine at residue 33 in our sequences replaced by arginine) and one in the hinge domain (arginine at residue 164 replaced by lysine); both of these differences result in conservative substitutions. A phylogenetic tree constructed using our ER sequence was consistent with previously published trees and the evolutionary relationships among taxa (Fig 1B, [6,8,88]). .

3.3 *CYP19* gene analysis.

3.3.1. Exon-intron size and organization

Through interrogation of the amphioxus genome assembly and cloning of all the *B. floridae CYP19* exons and introns, we determined the complete sequence of the gene (GenBank Accession Number HQ010363). Like all other *CYP19* genes, the amphioxus *CYP19* has nine coding exons, and these are well conserved in size (Figure 2, Table 1). As previously reported for *CYP19a1* the predominant ovarian aromatase in goldfish [17] and zebrafish [89], the amphioxus *CYP19* gene most closely resembles the situation of the human gene in which the PII (ovarian) promoter and untranslated first exon are contiguous with and immediately upstream of the ATG site in exon II [90,91]. In contrast, *CYP19a2*, the predominant brain aromatase of teleostean fish, has an untranslated first exon farther upstream and, in this respect, resembles the human ortholog, in which multiple promoters/untranslated first exons located as far as -93 kb from the translation initiation site are alternatively spliced in a tissue-specific manner to a common site in exon II such that the aromatase protein synthesized is identical in all tissues

[90,91], suggesting that tissue-specific promoters were acquired sequentially during the course of evolution (ovary>brain>placenta). From the ATG in exon II, the amphioxus *CYP19* is approximately 7 kb, much smaller than the human *CYP19* (30 kb) or either zebrafish *CYP19a1* (15 kb) or *CYP19a2* (12 kb), due primarily to shorter introns (Figure 2; Table 1). Worth noting here, our experimentally determined intronic sequences, when aligned with the amphioxus whole genome database, had a number of indels and mismatches, most notably a 1300 bp insert in intron III and a 216 bp insert in intron IV at the junction with exon V.

3.3.2 Identification of putative cis-regulatory elements

Regulation of *CYP19* expression and promoter structure varies considerably among taxa. In contrast to teleost fish in which aromatase expression in brain and ovary is controlled by two distinct genes and promoters, the *CYP19* of humans, other mammals and birds is a single gene with multiple promoters (also see section 3.3.1 above, and legend to Fig. 2). From genomic DNA, we amplified, cloned and sequenced 5'-flanking sequence 1184 bp upstream of the ATG in exon II (Genbank Accession Number HQ010363), which includes a TATA box at –187. Although overall sequence identity in the 5'-flanking region of the different *CYP19* genes was low, statistically over-represented motifs corresponding to known cis elements were identifiable. TRANSFAC analysis of the *B. floridae* 5'-flanking sequence revealed at least six potential transcription factor binding sites, each of which have been identified within the aromatase promoter from other taxa (Table 2). Notably, some forms of aromatase from other taxa (e.g., *CYP19a1b* expressed predominantly in teleostean brain [19,92] and several human tissue-specific *CYP19* promoters [91,93]) can be induced by estradiol exposure through direct ER interactions with estrogen-responsive elements (ERE) or indirect ER interactions with other

505 transcription factors and binding sites. A typical ERE consists of two hexameric half-sites
506 (AGGTCA) in opposite orientation (inverted repeats), separated by three nucleotides [94]. In
507 addition, several nuclear receptors, including ERα, ERRα and SF-1 can bind to ERE half-sites or
508 extended half-sites (TCAAGGTCA, also called ERREs or SFREs) [95]. While we did not
509 identify EREs upstream of the *B. floridae CYP19*, three largely conserved putative ERE half-
510 sites were found within the amphioxus *CYP19* promoter (designated by MATCH as ERR and
511 SF-1 binding sites, Table 2). Availability of a putative promoter of the amphioxus CYP19
512 provides an entry point for studying transcriptional regulation at this key phyletic level.
513
514 **3.4. Alignment of deduced amino acid sequences in conserved functional domains of**
515 **amphioxus and vertebrate aromatases**
516 Key functional domains of our deduced amino acid sequence were aligned with reported
517 *CYP19* sequences from the congener *B. belcheri* and representative vertebrates (Fig. 3).
518 Boundaries of conserved functional domains are as described by Simpson et al. [11] and
519 correspond to the following residues: human (I-helix 294-324, aromatic region 376-398, heme-
520 binding 424-443), *B. floridae* (I-helix 327-357, aromatic region 399-430, heme-binding 460-
521 475). Comparison among taxa revealed moderate conservation between amphioxus and
522 vertebrate sequences. Relative to *B. belcheri,* our sequence contained one difference in the
523 aromatase-specific conserved region, and one in the heme-binding region. Compared with the
524 human sequence, the *B. floridae* sequence exhibited 54% identity (17/31 residues) in the I-helical
525 domain, 56% (18/32 residues) in the aromatic region, and 56% (10/18 residues) in the heme-
526 binding domain. Within these regions, the four residues shown to contact the substrate by the
527 human aromatase (A306, D309, T310, F427; Fig. 4b, 4d) are also predicted to contact the

528 substrate by the amphioxus aromatase (A339, D342, T343, F463; Fig. 4a, 4c, Section 3.5). These
529 four residues are perfectly conserved among all taxa shown.
530
531 **3.5. Aromatase modeling, mapping and docking**
532 Using MODELLER, the human and amphioxus aromatase sequences are 40% identical
533 overall and, considering conservative mutations, show similarity for 60% of the amino acid
534 residues. In addition, the identical and similar residues are distributed evenly along the sequence,
535 and there are only 14 residues in gap regions for the sequence of 452 amino acids. Based on this
536 high level of sequence conservation, it is expected that a useful model of the amphioxus
537 aromatase can be constructed based on the structure of the human protein.
538 Figure 4A shows the amino acid residues in the binding site of the resulting amphioxus
539 aromatase model and the position and orientation of androstenedione (shown in grey) obtained
540 by docking. We note that the docking is predicted to be fairly accurate. Indeed, the 100
541 independent docking runs yielded docked androstenedione poses that can all be confined to a
542 cluster with a mean root mean square deviation (RMSD) of less than 0.8 Å, and all docked
543 structures have very similar interactions with the surrounding residues. In order to further test
544 the docking algorithm, we also docked androstenedione to the known structure of the human
545 aromatase (Figure 4B). The docked poses from the 100 docking runs formed a cluster with the
546 RMSD of less than 1.2 Å, and the lowest energy docked pose (grey) had an RMSD of less than 1
547 Å from the androstenedione pose in the X-ray structure (shown in violet). In addition to the
548 similar binding modes, the binding energies obtained in the two docking experiments (-10.7 and
549 -11.3 kcal/mol for human and amphioxus aromatase, respectively) suggest that androstenedione
550 is likely to bind to the human and amphioxus proteins with similar affinity.

Figure 4C and 4D show the percentage of nonbonded interactions between the small molecular probes from the computational solvent mapping and the amino residues in the human and amphioxus aromatase, respectively. We consider only the binding site residues within 6 Å from any androstenedione atom. The two fingerprints confirm the conservative character of the mutations in the binding site, and explain why the binding modes of androstenedione are so similar in the two proteins. The site includes 21 amino acid residues that have more than 1% of the nonbonded interaction contacts in one or both structures, but only one of these residues is mutated (from L372 to F404). In addition, as shown for F404 in Figure 4A and for L372 in Figure 4B, these residues interact with the bound androstenedione using backbone atoms rather than their side chains, and hence do not affect the binding features. Thus, all residues that are critical for the binding of small molecules are also highly conserved during the course of evolution. The conservation is not as strong for the less important residues: among the five positions in the binding site that have less than 1% of the nonbonded interaction contacts, two are mutated during the course of evolution (I305 to V338 and A307 to G340).

Based on the results described above, there is a remarkable degree of conservation in the predicted structure of the amphioxus and human aromatase proteins despite the approximately 500 million years of divergence between the cephalochordate and vertebrate lineages. While the overall amino acid identity is moderate (40%), binding site residues are highly conserved, and docking results indicate that androstenedione is likely to react within the catalytic site of the amphioxus protein as it does with human aromatase. In this regard, it would be of interest to compare the substrate affinity and catalytic activity of the two aromatase enzymes in the same membrane context.

3.6 Functional characterization of amphioxus ER

Paris et al. [6] inferred from limited proteolysis assays that bisphenol A binds the amphioxus ER but other classic ER ligands (estradiol, 3b-androstane-diol, 4-hydroxytamoxifen, diethylstilbestrol, enterolactone, ICI-182780) do not. The limited proteolysis assay indicates the ability of a compound to induce a conformational change in a protein that protects it from typsin digestion, as is generally observed upon binding of estrogens to the vertebrate ER LBD [6,96]. In this report, we quantified specific binding of radiolabeled estradiol to the human and amphioxus estrogen receptors as a more direct measurement of binding. When expressed *in vitro*, human ERα specifically bound tritiated estradiol in a saturable manner with high affinity (Fig. 5, Kd = 0.23 ± .046 nM). In contrast, no specific binding of estradiol to the amphioxus ER was detected in this assay (Fig. 5A).

When human ERα and amphioxus ER were transiently transfected into COS-7 cells, they produced proteins of the expected size (59 kD amphioxus, 66 kD human) with a similar efficiency. As expected, estradiol, bisphenol A, diethylstilbestrol and genistein activated human ERα, and 4-hydroxytamoxifen (an ER antagonist) did not activate human ERα. As shown in Fig. 5B, activation of human ERα by the weak estrogens bisphenol A and genistein was more variable (larger error bars), although this variability was not consistently observed. The amphioxus ER showed no constitutive activity beyond that of an empty expression vector. Transactivation by the amphioxus ER was not increased in the presence of estradiol or the other estrogenic compounds tested (Fig. 5B). These results are consistent with previous studies showing that amphioxus ER is not activated by ligands for the vertebrate ER [6,8,60]. Indeed a ligand for amphioxus ER has not been identified, although it has been demonstrated that amphioxus ER can serve as a competitive repressor for the hormone-activated SR [8,60].

3.7 Tissue distribution of aromatase and ER mRNA

Semi-quantitative RT-PCR was conducted to examine the expression of aromatase and ER transcripts in different amphioxus body segments (Fig 6A). As previously reported for aromatase enzyme activity [26], aromatase mRNA expression was limited to central (gonad-containing) segments, and expression was somewhat higher in females. Although ER mRNA was detectable in all three regions, the relative band intensity was tissue-related: expression was highest in gonad-containing segments (ovary > testis) and lower but approximately equal in cephalic and caudal segments (Fig 6B). Overall, these expression patterns are consistent with results from Bridgham et al [8], who used *in situ* hybridization to demonstrate that ER and SR are primarily expressed in gonads: ER and SR were co-expressed in oocytes, but in testes SR was broadly expressed and ER expression was more restricted.

4. CONCLUSIONS AND FUTURE PERSPECTIVES

The basic requirements of a functional chemical signaling system are (a) a messenger molecule; (b) a cellular receptor for recognition and signal transduction; and (c) a biological response. Results presented here reinforce the view that the cephalochordate amphioxus has the ability to synthesize estrogen, and also has the core molecular elements of a classical vertebrate ER-mediated signal transduction pathway. While modeling and docking studies predict that amphioxus aromatase will bind androgen, the substrate affinity, catalytic activity and other reaction properties of this enzyme remain to be evaluated. In addition, functional differences between vertebrate and amphioxus ERs and SRs indicate that mechanistic differences in estrogen

signaling must exist between the two groups. Indeed, evidence that aromatizable substrate is available and that estrogen is actually recognized as a chemical messenger that activates a cellular response in a biologically relevant context remains to be established.

What is clear from our new analyses of the amphioxus *CYP19* gene and aromatase protein is the remarkable degree of structural and functional conservation from amphioxus to humans. To place this in an evolutionary timeframe, the ancestral chordate represented by the common ancestor to contemporary vertebrates, amphioxus, and tunicates is estimated to have emerged 500 million years ago (Cambrian era). In view of this ancient history, it is surprising that a recognizable ancestral *CYP19* has not yet been found among the *CYP* genes in invertebrates. Although the possibility that *CYP19* was secondarily lost in invertebrates cannot be ruled out, a renewed search using the larval forms of invertebrates and a wider range of species could be productive in illuminating the evolution of this important member of the *CYP* family of genes.

In itself, conservation of a character, such as the ability to synthesize estrogen, signifies an important adaptive value. Moreover, coexpression of aromatase and ER in the gonads suggests a functional interaction, perhaps a paracrine/autocrine signaling role in regulating seasonal or cyclical gonadal growth as occurs in vertebrates. How can this be accomplished if, as we show here, amphioxus ER does not bind estradiol? One explanation is that estradiol is not a surrogate for the actual amphioxus estrogen. Certainly, many natural steroidal chemicals (estrone, estriol, catechol estrogens) have estrogenic or antiestrogenic bioactivity but differ substantially in their binding properties and spectrum of bioactivities when compared to estradiol, even when tested with mammalian ER. It is worth noting here that aromatization of androgen to estrogen occurs in three hydroxylation steps and accumulation of intermediates such as 19-nortestoserone is substantial with some aromatases (e.g., porcine blastocyst isoform; [97]).

643 To our knowledge, these steroids have not been tested with amphioxus ER although 19-
644 nortestosterone is reported to bind to the mammalian ERβ [98]. Additionally, estrone and
645 estradiol can be further metabolized to a variety of hydroxylated forms (e.g., at C2, C4).
646 Although these estrogens generally do not interact to any extent with mammalian ER, they
647 cannot be ruled out as ligands of the amphioxus ER.

648 Another way to explain discordance between estrogen synthesis and estrogen action is
649 that the early estrogen signaling system involved ER indirectly, for example, through
650 heterodimerization with another estrogen-activated nuclear receptor (ERR, SR), or through
651 binding with a different class of membrane-associated receptors (GPR30). These, in turn, could
652 activate ER through phosphorylation or other post-translational modification. Additionally, ERs
653 partner in protein-protein interactions with other nuclear factors by which they are tethered to
654 DNA binding motifs including Sp-1 and AP-1 recognition elements (Safe & Kim Journal of
655 Molecular Endocrinology (2008) 41, 263–275). Without testing a variety of reporter constructs,
656 it would be premature to conclude that the amphioxus estrogen/ER complex lacks
657 transcactivational activity.

658 If it can be proven that the role of ER in estrogen signaling in amphioxus is indirect, then
659 it is reasonable to postulate that direct estrogen binding/transactivation of ER is a feature that
660 was acquired secondarily during the course of evolution, concomitant with the ever-increasing
661 complexity of vertebrate organisms. This theory could explain the remarkable diversity and
662 complexity of estrogen signaling pathways in contemporary mammals: genomic/transcriptional;
663 rapid non-genomic/membrane-mediated; ligand- and ERE-dependent and independent (see
664 Introduction).The value of an evolutionary perspective is that it provides a conceptual framework
665 for organizing and analyzing information, thereby revealing common themes, unanswered

666 questions and new hypotheses for testing. At this point we cannot rule out the possibility that
667 endogenously synthesized estrogen is just a metabolic byproduct, or that the ER of adult
668 amphioxus is preadaptive or degenerate. The information provided here provides an entry point
669 for new molecular analysis. A key remaining challenge, however, is to demonstrate that estrogen
670 has biologically relevant effects at this phyletic level.
671
672 **ACKNOWLEDGEMENTS**
673
674 Supported by grants from the NIEHS P42 ES07381 (GVC, SV) and EPA (STAR-RD831301)
675 (GVC), a Ruth L Kirschstein National Research Service Award (AT, F32 ES013092-01), an NIH
676 traineeship (SS, SG), a NATO Fellowship (AN) and the Boston University Undergraduate
677 Research Program (LC). The human ERαand 3xERE-TATA-luc reporter plasmids were
678 generously provided by Dr. Donald McDonnell.
679
680

681 **LEGENDS**

682

683 **Figure 1: Phylogenetic trees of (A) aromatase and (B) ER proteins.** Trees were
684 constructed to demonstrate the phylogenetic position of our amphioxus aromatase and ER
685 sequences**;** topologies were consistent with previously published trees [5,7,31] and the
686 evolutionary relationships among taxa [22]. GenBank Accession numbers are given
687 parenthetically. (A) Deduced amino acid sequence of amphioxus aromatase was aligned with
688 vertebrate *CYP19* sequences and other representative *CYP* sequences. The maximum likelihood
689 tree was rooted with the human *CYP17* and *CYP21* sequences (CYP Clan 2). Accession
690 numbers: Amphioxus *CYP19* (ABA47317.1) zebrafish *CYP19a1/A* (AA65788.1), zebrafish
691 *CYP19a2/B* (AAK00642.1), killifish *CYP19a1/A* (AAR97268.1), killifish *CYP19a2/B*
692 (AAR97269.1), Human *CYP19* (NP_112503.1), mouse *CYP19* (P28649.1), Human *CYP17*
693 (AAA36405; Human *CYP21* (NP_000491). Numbers indicate percentage of 100 bootstrap
694 replicates supporting each node. (B) Deduced amino acid sequence of amphioxus ER was
695 aligned with vertebrate ER and ERR sequences, and a Neighbor-Joining tree was constructed.
696 Protostomes (mollusc and annelid) ER sequences were not included in this analysis (see [10] for
697 a thorough analysis of the evolutionary position of these genes). Amphioxus *ER* (EF 554313.1),
698 teleost *ERβb* (zebrafish, NP_777287; goldfish, Q9IAL9), teleost *ERβa* (zebrafish, NP_851297;
699 trout, CAC06714; goldfish, Q9W669; medaka, AAX14000; Salmon AAR92486), mammal *ERβ*
700 (human, CAA67555; rat, U57439; mouse, AAB51132), Lamprey *ER* (AAK20929), teleost
701 *ERα* (Medaka, P50241; Salmon, P50242; Trout, P16058; Goldfish *ER*, AAL12298; zebrafish,
702 NP_694491; mammal *ERα* (Mouse, NP_031982; Rat, P06211), *ERRs* (Amphioxus *ERR*
703 AAU88062; Human *ERRα* (NP_004442), ERRβ (O95718), and *ERRγ* (AAQ93381). Numbers

on nodes indicate percentage of 1000 bootstrap replicates supporting each node, and triangles indicate nodes collapsed for simplicity.

Figure 2. Comparison of *CYP19* genes in amphioxus, human and zebrafish. The genomic organization of the coding region of the single copy *CYP19* gene in the human (A, upper panel; NM 000103.3; [108] is compared to that of amphioxus (HQ010363) and zebrafish *CYP19a1(A)* and *CYP19a2(B)* genes (NM 131154.2 and NM131642.1 [109]) (B, lower panel). Exons II – X are labeled in human (panel A) and correspondingly color coded in amphioxus and zebrafish (panel B). The translation initiation (*) and the stop (°) codons are indicated. Note that the ovarian promoter/untranslated first exon of the human *CYP19* (PII) is contiguous with exon II, whereas the placental (I.1) and brain (I.f) promoters and first exons are located ~93 kb and 33 kb upstream of the ATG in exon II. The untranslated first exon in amphioxus and zebrafish *CYP19A1(A)* is contiguous with exon II, while the untranslated first exon of zebrafish *CYP19A2(B),* like that of human I.f is further upstream. Also, the very long exon X (3'-untranslated region) of zebrafish *CYP19A2(B)* has a ~250-bp region deleted from the mRNA.

Figure 3: **Sequence alignment of conserved functional domains of aromatases in amphioxus and representative vertebrates.** Boundaries are as described by Simpson et al. [11] for amino acid residues of the human aromatase: I-helix 294-324: aromatic region 376-398: heme-binding domain 426-443). Identical and similar amino acid residues are marked by asterisks and dots, respectively. GenBank accession numbers given in Figure 1 legend or as follows: amphioxus (*B. floridae*, EF554313.1; *B. belcheri*, BAF61105.1), Dogfish (ABB53418.1), Killifish

(*CYP19A1(A)*: AAR97268, *CYP19A2(B*): AAR97269), Xenopus (BAA90529), Zebra finch (AAB32404.1), Turtle (AAG09376), Rat (NP_036885.1), Pig (AAB51387). .

Figure 4: Homology model of androstenedione docked within the active site of aromatase in (A) amphioxus and (B) human and plot of non-bonded interactions in (C) amphioxus and (D) human. See sections 2.3 and 3.5 for detailed methods and results describing evolutionarily conserved residues. Also, compare with conserved residues identified by sequence alignment (Fig. 3; section 3.3.2).

Figure 5: Functional comparison of amphioxus ER through binding experiments and cell-based reporter assays. Amphioxus estrogen receptor (BfER) and human estrogen receptor alpha (HsERα) constructs containing a V5 epitope tag were expressed in rabbit reticulocyte (A) and in COS-7 cells (A inset, B). (A) Tritiated estradiol was specifically bound by the human ERα (triangles) as expected, but not by the amphioxus ER (squares). Representative results from one of four independent experiments are shown. (inset) A western blot with a v5 antibody showing expression of BfER and ERα transfected into COS-7 cells. (B) Amphioxus ER (white bars), human ERα (grey bars), or an empty expression vector (pcDNA, black bars) were transfected into COS-7 cells along with a luciferase reporter driven by three estrogen responsive elements (see methods). The y-axis shows the ratio of luminescence by the luciferase reporter to luminescence by a transfection control reporter. Units are normalized such that the value for the DMSO-treated empty expression vector is equal to one.

Figure 6: Tissue-specific expression of (A) *CYP19* and (B) ER mRNAs in amphioxus, as determined by semiquantitative RT-PCR analysis. Tissues were collected during the period of reproductive activity. H, head; T, testis; O, ovary; Ta, tail; C, control (no template). Products were separated on 1% agarose gels in 0.5X TBE, stained with ethidium bromide and visualized under ultraviolet light.

REFERENCES CITED

[1] Engel, L. (1974) The biosynthesis of estrogen. In: Greep RO (ed) Handbook of Physiology vol 2, sect 7, part I, American Physiological Society, Washington DC, pp. 467-83.

[2] Santen, R., Brodie, H., Simpson, E., Siiteri, P. and Brodie, A. (2009) History of aromatase: saga of an important biological mediator and therapeutic target. Endocrine Reviews 30, 343-75.

[3] Heldring, N., Pike, A., Andersson, S., Matthews, J., Cheng, G., Hartman, J., Tujague, M., Strom, A., Treuter, E., Warner, M. and Gustafsson, J.-A. (2007) Estrogen receptors: how do they signal and what are their targets. Physiol Rev 87, 905-31.

[4] Castro, L., Santos, M. and Reis-Henriques, M. (2005) The genomic environment around the Aromatase gene: evolutionary insights. BMC Evolutionary Biology 5, 43.

[5] Mizuta, T. and Kubokawa, K. (2007) Presence of sex steroids and cytochrome P450 genes in amphioxus. Endocrinology 148, 3554-65.

[6] Paris, M., Pettersson, K., Schubert, M., Bertrand, S., Pongratz, I., Escriva, H. and Laudet, V. (2008) An amphioxus orthologue of the estrogen receptor that does not bind estradiol: Insights into the estrogen receptor evolution. BMC Evolutionary Biology 8, 219.

[7] Schubert, M., Brunet, F., Paris, M., Bertrand, S., Benoit, G. and Laudet, V. (2008) Nuclear receptor signaling in amphioxus. Dev Genes Evol 218, 651-65.

[8] Bridgham, J., Brown, J., Rodrigues-Mari, A., Catchen, J. and Thornton, J. (2008) Evolution of a new function by degenerative mutation in cephalochordate steroid receptors. PLoS Genet 4, e1000191.

[9] Keay, J., Bridgham, J. and Thornton, J. (2006) The *Octopus vulgaris* estrogen receptor is a constitutive transcriptional activator: evolutionary and functional implications. Endocrinology 147, 3861-9.

[10] Keay, J. and Thornton, J. (2009) Hormone-activated estrogen receptors in annelid invertebrates: implications for evolution and endocrine disruption. Endocrinology 150, 1731-8.

[11] Simpson, E., Mahendroo, M., Means, G., Kilgore, M., Hinshelwood, M., Graham-Lorence, S., Amarneh, B., Ito, Y., Fisher, C., MD, M., Mendelson, C. and Bulun, S. (1994) Aromatase cytochrome P450, the enzyme responsible for estrogen biosynthesis. Endocrine Reviews 15, 342-55.

[12] Amarneh, B., Corbin, C., Peterson, J., Simpson, E. and Graham-Lorence, S. (1993) Functional domains of human aromatase cytochrome P450 characterized by linear alignment and site-direct mutagenesis. Mol Endocrinol 7, 1617-24.
[13] Shimozawa, O., Sakaguchi, M., Ogawa, H., Harada, N., Mihara, K. and Omura, T. (1993) Core glycosylation of cytochrome P-450(arom). Evidence for localization of N terminus of microsomal cytochrome P-450 in the lumen. J Biol Chem 268, 21399-402.
[14] Ghosh, D., Griswold, J., Erman, M. and Pangborn, W. (2009) Structural basis for androgen specificity and oestrogen sythesis in human aromatase. Nature 457, 219-23.
[15] Ghosh, D., Griswold, J., Erman, M. and Pangborn, W. (2010) X-ray structure of human aromatase reveals an androgen-specific active site. Journal of Steroid Biochemistry and Molecular Biology 119, 197-202.
[16] Graham-Lorence, S., Khalil, M., Lorence, M., Mendelson, C. and Simpson, E. (1991) Structure-function relationships of human aromatase cytochrome P-450 using molecular modeling and site-directed mutagenesis. J Biol Chem 266, 11939-46.
[17] Callard, G.V. and Tchoudakova, A. (1997) Evolutionary and functional significance of two CYP19 genes differentially expessed in brain and ovary of goldfish. J Steroid Biochem Mol Biol 61, 387-92.
[18] Kishida, M. and Callard, G.V. (2001) Distinct cytochrome P450 aromatase isoforms in zebrafish (Danio rerio) brain and ovary are differentially programmed and estrogen regulated during early development. Endocrinology 142, 740-50.
[19] Tchoudakova, A., Kishida, M., Wood, E. and Callard, G. (2001) Promoter characteristics of two cyp19 genes differentially expressed in the brain and ovary of teleost fish. J. Steroid Biochem. Mol. Biol. 78, 427-439.
[20] Conley, A., Corbin, C. and Hughes, A. (2009) Adaptive evolution of mammalian aromatases: lessons from Suiformes. Journal of Experimental Zoology Part A: Ecological Genetics and Physiology 311A, 346-57.
[21] Corbin, C., Hughes, J., Heffelfinger, T., Berger, T., Waltzek, J., Roser, T., Santos, M., Miglino, M., Oliveira, F., Braga, F., Meirelles, F. and Conley, A. (2007) Evolution of suiform aromatases: ancestral duplication with conservation of tissue-specific expression in the collared peccary (*Pecari tayassu*). Journal of Molecular Evolution 65, 403-12.
[22] Chiang, E.F., Yan, Y.L., Guiguen, Y., Postlethwait, J. and Chung, B. (2001) Two Cyp19 (P450 aromatase) genes on duplicated zebrafish chromosomes are expressed in ovary or brain. Mol Biol Evol 18, 542-50.
[23] Kishida, M., McLellan, M., Miranda, J. and Callard, G.V. (2001) Estrogen and xenoestrogens upregulate the brain aromatase isoform (P450aromB) and perturb markers of early development in zebrafish (*Danio rerio*). Comp Biochem Physiol B Biochem Mol Biol 129, 261-8.
[24] Zhao, J., Mak, P., Tchoudakova, A., Callard, G. and Chen, S. (2001) Different catalytic properties and inhibitor responses of the goldfish brain and ovary aromatase isozymes. Gen Comp Endocrinol 123, 180-91.
[25] Tong, S., Chiang, E.F., Hsiao, P. and Chung, B. (2001) Phylogeny, expression and enzyme activity of zebrafish cyp19 (P450 aromatase) genes. J Steroid Biochem Mol Biol 79, 299-303.
[26] Gelinas, D., Pitoc, G. and Callard, G. (1998) Isolation of a goldfish brain cytochrome P450 aromatase cDNA: mRNA expression during the seasonal cycle and after steroid treatment. Mol Cell Endocrinol 138, 81-93.

[27] Tchoudakova, A. and Callard, G.V. (1998) Identification of multiple CYP19 genes encoding different cytochrome P450 aromatase isozymes in brain and ovary. Endocrinology 139, 2179-89.
[28] Bulun, S., Sebastian, S., Takayama, K., Suzuki, T., Sasano, H. and Shozu, M. (2003) The human CYP19 (aromatase P450) gene: update on physiologic roles and genomic organization of promoters. J Steroid Biochem Mol Biol 86, 219-24.
[29] Wang, H., Li, R. and Hu, Y. (2009) The alternative non-coding exons 1 of aromatase (Cyp19) gene modulate gene expression in a post-transcriptional manner. Endocrinology 150, 3301-7.
[30] Nelson, D. (1998) Metazoan cytochrome P450 evolution. Comp BIochem Physiol C Toxicol Pharmacol 121, 15-22.
[31] Markov, G., Tavares, R., Dauphin-Villemant, C., Demeneix, B., Baker, M. and Laudet, V. (2009) Independent elaboration of steroid hormone signaling pathways in metazoans. Proc Natl Acad Sci USA 106, 11913-8.
[32] Reitzel, A. and Tarrant, A. (2010) Correlated evolution of androgen receptor and aromatase revisited. Mol Biol Evol 27, 2211-5.
[33] Holland, L., Albalat, R., Azumi, K., Benito-Gutierrez, E., Blow, M., Bronner-Fraser, M., Brunet, F., Butts, T., Candiani, S., Dishaw, L., Ferrier, D., Garcia-Fernandez, J., Gibson-Brown, J., Gissi, C., Godzik, A., Hallbook, F., Hirose, D., Hosomichi, K., Ikuta, T., Inoko, H., Kasahara, M., Kasamatsu, J., Kawashima, T., Kimura, A., Kobayashi, M., Kozmik, Z., Kubokawa, K., Laudet, V., Litman, G., McHardy, A., Meulemans, D., Nonaka, M., Olinski, R., Pancer, Z., Pennacchio, L., Pestarino, M., Rast, J., Rigoutsos, I., Robinson-Rechavi, M., Roch, G., Saiga, H., Sasakura, Y., Satake, M., Satou, Y., Schubert, M., Sherwood, N., Shiina, T., Takatori, N., Tello, J., Vopalensky, P., Wada, S., Xu, A., Ye, Y., Yoshida, K., Yoshizaki, F., Yu, J., Zhang, Q., Zmasek, C., de Jong, P., Osoegawa, K., Putnam, N., Rokhsar, D., Satoh, N. and Holland, P. (2008) The amphioxus genome illuminate vertebrate origins and cephalochordate biology. Genome Res 18, 1100-1111.
[34] Callard, G., Pudney, J., Kendall, S. and Reinboth, R. (1984) In vitro conversion of androgens to estrogen in amphioxus gonadal tissues. General and Comparative Endocrinology 56, 53-58.
[35] Mizuta, T., Asahina, K., Suzuki, M. and Kubokawa, K. (2008) In vitro conversion of sex steroids and expression of sex steroidogenic enzyme genes in amphioxus ovary. J Exp Zool 309A, 83-93.
[36] Green, S., Walter, P., Kumar, V., Krust, A., Bornert, J., Argos, P. and Chambon, P. (1986) Human estrogen receptor cDNA: sequence, expression and homology to v-erbA. Nature 320, 134-9.
[37] Kuiper, G., Enmark, E., Pelto-Huikko, M., Nilsson, S. and Gustafsson, J.-A. (1996) Cloning of a novel estrogen receptor expressed in rat prostate and ovary. Proc Natl Acad Sci U S A 93, 5925-30.
[38] Thornton, J. (2001) Evolution of vertebrate steroid receptors from an ancestral estrogen receptor by ligand exploitation and serial genome expansions. Proc Natl Acad Sci U S A 98, 5671-6.
[39] Kumar, V., Green, S., Stack, G., Berry, M., Jin, J. and Chambon, P. (1987) Functional domains of the human estrogen receptor. Cell 51, 941-51.

[40] Tora, L., White, J., Brou, C., Tasset, D., Webster, N., Scheer, E. and Chambon, P. (1989) The human estrogen receptor has two independent non acidic transcriptional activation functions. Cell 59, 477-87.
[41] Zwart, W., de Leeuw, R., Rondaij, M., Neefjes, J., Mancini, M. and Michalides, R. (2010) The hinge region of the human estrogen receptor determines functional synergy between AF-1 and AF2 in the quantitative response to estradiol and tamoxifen. Journal of Cell Science 123, 1253-61.
[42] Pratt, W. and Toft, D. (1997) Steroid receptor interactions with heat shock protein and immunophilin chaperones. Endocrine Reviews 18, 306-30.
[43] Powell, E. and Xu, W. (2008) Intermolecular interactions identify ligand-selective activity of estrogen receptor α/β dimers. Proc Natl Acad Sci U S A 105, 19012-7.
[44] Kos, M., Denger, S., Reid, G. and Gannon, F. (2002) Upstream open reading frames regulate the translation of the multiple mRNA variants of the estrogen receptor alpha. J Biol Chem 277, 37131-8.
[45] Poola, I., Koduri, S., Chatra, S. and Clarke, R. (2000) Identification of twenty alternatively spliced estrogen receptor alpha mRNAs in breast cancer cell lines and tumors using splice targeted primer approach. Journal of Steroid Biochemistry and Molecular Biology 72, 249-58.
[46] Poola, I. and Speirs, V. (2001) Expression of alternatively spliced estrogen receptor alpha mRNAs is increased in breast cancer tissues. Journal of Steroid Biochemistry and Molecular Biology 78, 459-69.
[47] Herynk, M.H. and Fuqua, S.A. (2004) Estrogen receptor mutations in human disease. Endocr Rev 25, 869-98.
[48] Hirata, S., Shoda, T., Kato, J. and Hoshi, K. (2003) Isoform/variant mRNAs for sex steroid hormone receptors in humans. Trends Endocrinol Metab 14, 124-9.
[49] Poola, I., Abraham, J. and Baldwin, K. (2002) Identification of ten exon deleted ERbeta mRNAs in human ovary, breast, uterus and bond tissues: alternative splicing pattern of estrogen receptor beta mRNA is distinct from that of estrogen recpetor alpha. FEBS Letters 516, 133-8.
[50] Losel, R. and Wehling (2003) Nongenomic actions of steroid hormones. Nat Rec Mol Cell Biol 4, 46-56.
[51] Wendler, A., Baldi, E., Harvey, B., Nadal, A., Norman, A. and Wehling, M. (2010) Position paper: Rapid responses to steroids: Current status and future prospects. Eur J Endocrinol 162, 825-30.
[52] Levin, E. (2009) Membrane oestrogen receptor alpha signalling to cell functions. Journal of Physiology 587, 5019-23.
[53] Pedram, A., Razandi, M., Sainson, R., Kim, J., Hughes, C. and Levin, E. (2007) A conserved mechanism for steroid receptor translocation to the plasma membrane. Journal of Biological Chemistry 282, 22278-88.
[54] Pedram, A., Razandi, M. and Levin, E. (2006) Nature of functional estrogen receptors at the plasma membrane. Molecular Endocrinology 20, 1996-2009.
[55] Thomas, P., Pang, Y., FIlardo, E. and Dong, J. (2005) Identity of an estrogen membrane receptor coupled to a G protein in human breast cancer cells. Endocrinology 146, 624-32.
[56] Thomas, P., Alyea, R., Pang, Y., Peyton, C., Dong, J. and Berg, A. (2010) Conserved estrogen binding and signaling functions of the G protein-coupled estrogen receptor 1 (GPER) in mammals and fish. Steroids 75, 595-602.

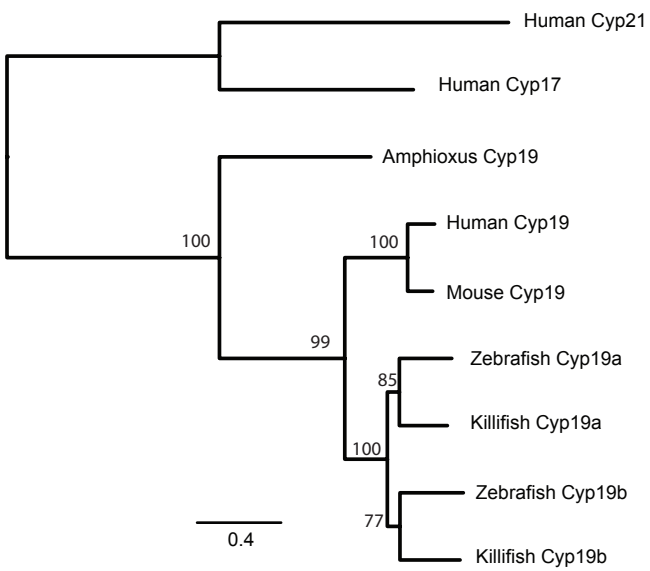
[57] Pang, Y. and Thomas, P. (2009) Involvement of estradiol-17beta and its membrane receptor, G protein coupled receptor 30 (GPR30) in regulation of oocyte maturation in zebrafish, Danio rario. Gen Comp Endocrinol 161, 58-61.
[58] Canesi, L., Ciacci, C., Betti, M., Lorusso, L., Marchi, B., Burattini, S., Falcieri, E. and Gallo, G. (2004) Rapid effects of 17β-estradiol on cell signaling and function of *Mytilus* hemocytes. Gen Comp Endocrinol 136, 58-71.
[59] Matsumoto, T., Nakamura, A., Mori, K., Akiyama, I., Hirose, H. and Takahashi, Y. (2007) Oyster estrogen receptor: cDNA cloning and immunolocalization. Gen Comp Endocrinol 151, 195-201.
[60] Katsu, Y., Kubokawa, K., Urushitani, H. and Iguchi, T. (2010) Estrogen-dependent transactivation of amphioxus steroid hormone receptor via both estrogen and androgen response elements. Endocrinology 151, 639-48.
[61] Thornton, J., Need, E. and Crews, D. (2003) Resurrecting the ancestral steroid receptor: ancient origin of estrogen signaling. Science 301, 1714-1717.
[62] Greytak, S., Tarrant, A., Hahn, M. and Callard, G. (2010) Estrogen responses in killifish (Fundulus heteroclitus) from polluted and unpolluted environments are site- and gene-specific. Aquatic Toxicology 99, 291-9.
[63] Greytak, S. and Callard, G. (2005) Isolation and characterization of two cytochrome P450 aromatase forms in killifish (*Fundulus heteroclitus*): differential expression in fish from polluted and unpolluted environments. Aquatic Toxicology 71, 371-89.
[64] Greytak, S. and Callard, G. (2007) Cloning of three estrogen receptors (ER) from killifish (Fundulus heteroclitus): Differences in populations from polluted and reference environments. Gen Comp Endocrinol 150, 174-88.
[65] Nelson, D. Cytochrome P450 Homepage. http://drnelson.uthsc.edu/cytochromeP450.html.
[66] Stamatakis, A., Hoover, P. and Rougemont, J. (2008) A rapid bootstrap algorithm for the RAxML Web-Servers. Systematic Biology 75, 758-71.
[67] Abascal, F., Zardoya, R. and Posada, D. (2005) ProtTest: selection of best-fit models of protein evolution. Bioinformatics 21, 2104-5.
[68] Felsenstein, J. (2004) PHYLIP (Phylogeny Inference Package) version 3.64, Department of Genome Sciences and Department of Biology, University of Washington, Seattle (distributed by the author; http://evolution.genetics.washington.edu/phylip.html).
[69] Kel, A., Gossling, E., Reuter, I., Cheremushkin, E., Kel-Margoulis, O. and Wingender, E. (2003) MATCH: A tool for searching transcription factor binding sites in DNA sequences. Nucleic Acids Res 31, 3576-9.
[70] Berman, H., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T., Weissig, H. and Shindyalov, I.B., PE (2000) The Protein Data Bank. Nucleic Acids Res 28.
[71] Eswar, N., Marti-Renom, M., Webb, M., Madhusudhan, D., Eramian, D., Shen, M., Pieper, U. and Sali, A. (2006) Comparative protein structure modeling with MODELLER. Current Protocols in Bioinformatics, John Wiley & Sons, Inc., Supplement 15, 5.6.1-5.6.30.
[72] Marti-Renom, M., Stuart, A., Fiser, A., Sanchez, R., Melo, F. and Sali, A. (2000) Comparative protein structure modeling of genes and genomes. Annu Rev Biophys Biomol Struct 29, 291-325.

[73] Landon, M., Lancia, D., Jr, Yu, J., Thiel, S. and Vajda, S. (2007) Identification of hot spots within druggable sites of proteins by computational solvent mapping. J Med Chem 50, 1231-40.
[74] Brenke, R., Kozakov, D., Chuang, G.-Y., Beglov, D., Hall, D., Landon, M., Mattos, C. and Vajda, S. (2009) Fragment-based identification of druggable "hot spots" of proteins using Fourier domain correlation techniques. Bioinformatics 25, 621-7.
[75] Hajduk, P., Huth, J. and Fesik, S. (2005) Druggability indices for protein targets derived from NMR-based screening data. J Med Chem 48, 2518-25.
[76] Mattos, C. and Ringe, D. (1996) Locating and characterizing binding sites on proteins. Nat Biotech 14, 595-9.
[77] Landon, M., Lieberman, R., Hoang, Q., Ju, S., Caaveiro, H., Orwig, S.K., D, Brenke, R., Chuang, G.-Y., Beglov, D., Vajda, S., Petsko, G. and Ringe, D. (2009) Detection of ligand binding hot spots on protein surfaces via fragment-based methods: Application to DJ-1 and glucocerebrosidase. J Comp Aided Mol Des Jun 12.
[78] Sheu, S., Kaya, T., Waxman, D. and Vajda, S. (2005) Exploring the binding site structure of the PPAR-g ligand binding domain by computational solvent mapping. Biochemistry 41, 1193-1209.
[79] Landon , M., Amaro, R., Baron, R., Ngan, C., Ozonoff, D., McCammon, J. and Vajda, S. (2008) Novel druggable hot spots in avian influenza neuraminidase H5N1 revealed by computational solvent mapping of a reduced and representative receptor ensemble. Chem Biol Drug Des 71, 106-16.
[80] Ngan, C.-H., Beglov, D., Rudnitskay, A., Kozakov, D., Waxman, D. and Vajda, S. (2009) The structural basis of pregnane X receptor binding promiscuity. Biochemistry 48, 11572-81.
[81] McDonald, I. and Thornton, J. (1994) Satisfying hydrogen bonding potential in proteins. J Mol Biol 238, 777-93.
[82] Morris, G., Goodsell, D., Halliday, R., Huey, R., Hart, W., Belew, R. and Olson, A. (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. J Computational Chem 19, 1639-62.
[83] Wittmann, B., Sherk, A. and McDonnell, D. (2007) Definition of functionally important mechanistic differences among selective estrogen receptor down-regulators. Cancer Research 67, 9549-60.
[84] Hawkins, M.B. and Thomas, P. (2004) The unusual binding properties of the third distinct teleost estrogen receptor subtype ERba are accompanied by highly conserved amino acid changes in the ligand binding domain. Endocrinology 145, 2968-2977.
[85] Jensen, B. and Hahn, M. (2001) cDNA cloning and characterization of a high affinity aryl hydrocarbon receptor in a cetacean, the beluga *Delphinapterus leucas*. Toxicol Sci 64, 41-56.
[86] Karchner, S., Franks, D. and Hahn, M. (2005) AHR1B, a new functional aryl hydrocarbon receptor in zebrafish: tandem arrangement of *ahr1b* and *ahr2* genes. Biochemistry Journal 391, 153-161.
[87] Hall, J. and McDonnell, D. (1999) The estrogen receptor b-isoform (ERb) of the human estrogen receptor modulates ERa transcriptional activity and is a key regulatory of cellular response to estrogens and antiestrogens. Endocrinology 140, 5566-5578.

[88] Baker, M. and Chandsawangbhuwana, C. (2008) Motif analysis of amphioxus, lamprey and invertebrate estrogen receptors: Toward a better understanding of estrogen receptor evolution. Biochem Biophys Res Commun 371, 724-8.
[89] Kazeto, Y., Ijiri, S., Place, A.R., Zohar, Y. and Trant, J.M. (2001) The 5'-flanking regions of CYP19A1 and CYP19A2 in zebrafish. Biochem Biophys Res Commun 288, 503-8.
[90] Kamat, A., Hinshelwood, M., Murray, B. and Mendelson, C. (2002) Mechanism in tissue-specific regulation of estrogen biosynthesis in humans. Trends Endocrinol Metab 13, 122-8.
[91] Kumar, P., Kamat, A. and Mendelson, C. (2009) Estrogen receptor alpha (ERalpha) mediates stimulatory effects of estrogen on aromatase (CYP19) gene expression in human placenta. Mol Endocrinol 23, 784-93.
[92] Callard, G., Tchoudakova, A., Kishida, M. and Wood, E. (2001) Differential tissue distribution, developmental programming, estrogen regulation and promoter characteristics of cyp19 genes in teleost fish. J Steroid Biochem Mol Biol 79, 305-14.
[93] Chen, D., Reierstad, S., Lu, M., Lin, Z., Ishikawa, H. and Bulun, S.E. (2009) Regulation of breast cancer-associated aromatase promoters. Cancer Lett 273, 15-27.
[94] Glass, C. (1994) Differential recognition of target genes by nuclear receptor monomers, dimers and heterodimers. Endocrine Reviews 15, 391-407.
[95] Vanacker, J.-M., Petterson, K., Gustafsson, J.-A. and Laudet, V. (1999) Transcriptional activities of the orphan nuclear receptor ERR alpha (estrogen receptor-related receptor-alpha). Molecular Endocrinology 13, 764-773.
[96] Elliston, J. and Katzenellenbogen, B. (1988) Comparative analysis of estrogen receptors covalently labeled with an estrogen and an antiestrogen in several estrogen target cells as studied by limited proteolysis. Journal of Steroid Biochemistry 29, 559-69.
[97] Kao, Y., Higashiyama, T., Sun, X., Okubo, T., Yarborough, C., Choi, I., Osawa, Y., Simmen, F. and Chen, S. (2000) Catalytic differences between porcine blastocyst and placental aromatase isozymes. Eur J Biochem 267, 6134-9.
[98] Kuiper, G., Carlsson, B., Grandien, K., Enmark, E., Haggblad, J., Nilsson, S. and Gustafsson, J.-A. (1997) Comparison of the ligand binding specificity and transcript tissue distribution of estrogen receptors alpha and beta. Endocrinology 138, 863-70.
[99] Chen, S., Ye, J., Kijima, I., Kinoshita, Y. and Zhou, D. (2005) Positive and negative transcriptional regulation of aromatase expression in human breast cancer tissue. J Steroid Biochem Mol Biol 95, 17-23.
[100] Shozu, M., Zhao, Y., Bulun, S.E. and Simpson, E. (1998) Multiple splicing events involved in regulation of human aromatase expression by a novel promoter, 1.6. Endocrinology 139, 1610-7.
[101] Ramachandran, B., Schlinger, B., Arnold, A. and Campagnoni, A. (1999) Zebra finch aromatase gene expression is regulated in the brain through an alternate promoter. Gene 240, 209-16.
[102] Honda, S., Harada, N. and Takagi, Y. (1996) The alternative exons 1 of the mouse aromatase cytochrome P-450 gene Biochim et Biophysica Acta 1305, 145-50.
[103] Chow, J., Simpson, E. and Boon, W. (2009) Alternative 5'-untranslated first exons of the mouse *Cyp19A1* (aromatase) gene. Journal of Steroid Biochemistry and Molecular Biology 115, 115-25.
[104] Tong, S.-K. and Chung, B.-c. (2003) Analysis of zebrafish *cyp19* promoters. Journal of Steroid Biochemistry and Molecular Biology 86, 381-6.

[105] Fitzpatrick, S. and Richards, J. (1993) cis-acting elements of the rat aromatase promoter required for cyclic adenosine 3',5'-monophosphate induction in ovarian granulosa cells and constitutive expression in R2C Leydig cells. Mol Endocrinol 7, 341-54.
[106] Chen, S., Itoh, T., Wu, K., Zhou, D. and Yang, C. (2003) Transcriptional regulation of aromatase expression in human breast tissue. Journal of Steroid Biochemistry and Molecular Biology 83, 93-9.
[107] Zeitoun, K., Takayama, K., Michael, M. and Bulun, S.E. (1999) Stimulation of aromatase P450 (II) activity in endometriosis and its inhibition in endometrium are regulated by competitive binding of steroidogenic factor-1 and chicken ovalbumin upstream promoter transcription factor to the same cis-acting element. Mol Endocrinol 13, 239-53.
[108] Mahendroo, M., Means, G., Mendelson, C. and Simpson, E. (1991) Tissue-specific expression of human P-450AROM. The promoter responsible for expression in adipose tissue is different from that utilized in placenta. J Biol Chem 266, 11276-81.
[109] Gates, M., Kim, L., Egan, E., Cardozo, T., Sirotkin, H., Dougan, S., Lashkari, D., Abagyan, R., Schier, A. and Talbot, W. (1999) A genetic linkage map for zebrafish: comparative analysis and localization of genes and expressed sequences. Genome Res 9, 334-47.

A.


Human Cyp21
Human Cyp17
Amphioxus Cyp19
Human Cyp19
Mouse Cyp19
Zebrafish Cyp19a
Killifish Cyp19a
Zebrafish Cyp19b
Killifish Cyp19b
100
100
99
85
100
77
0.4

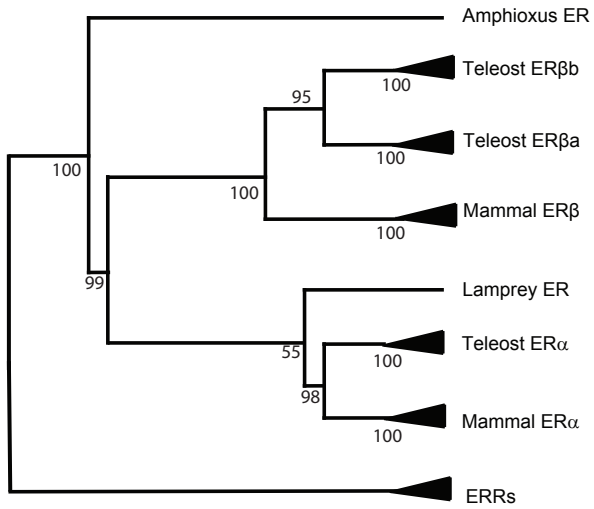
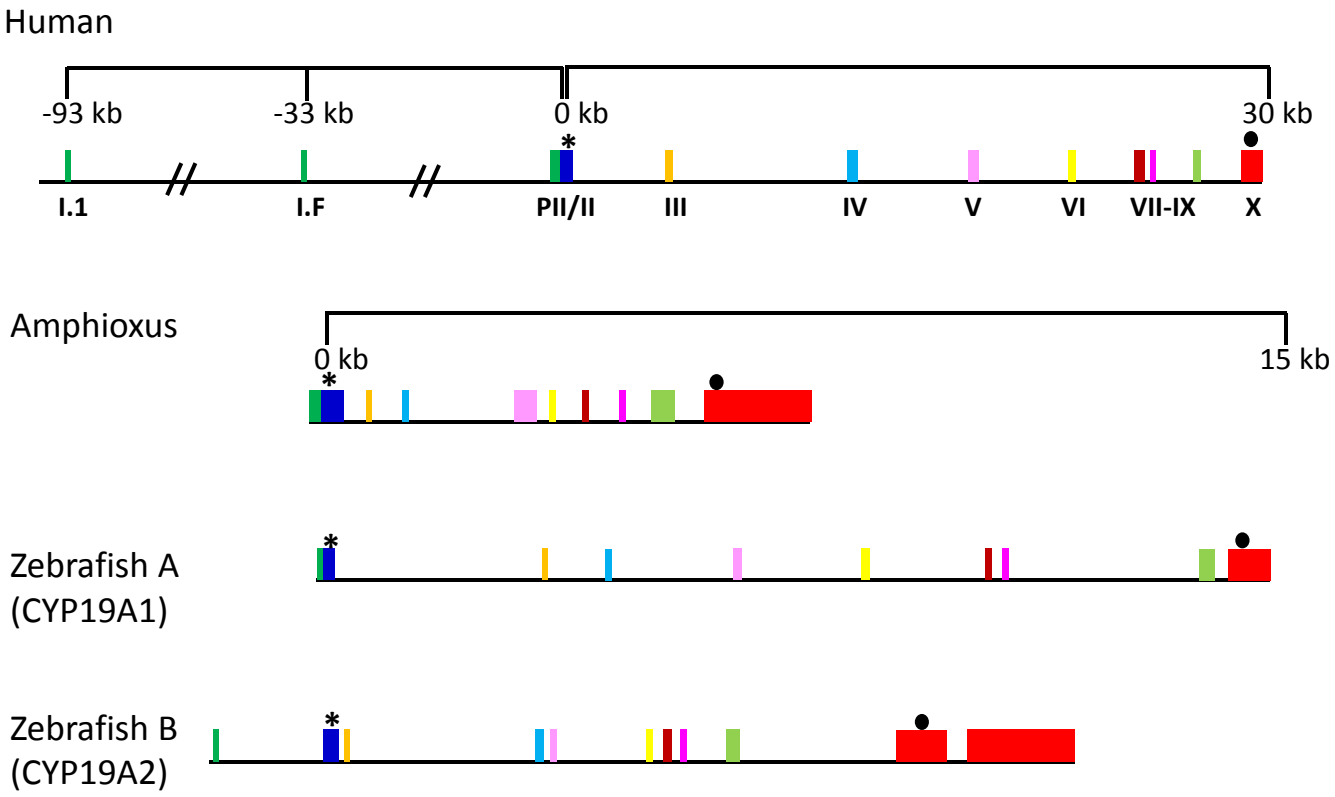

Amphioxus ER
Teleost ERβb
Teleost ERβa
Mammal ERβ
Lamprey ER
Teleost ERα
Mammal ERα
ERRs
100
95
100
100
100
99
55
100
98
100

Figure 2

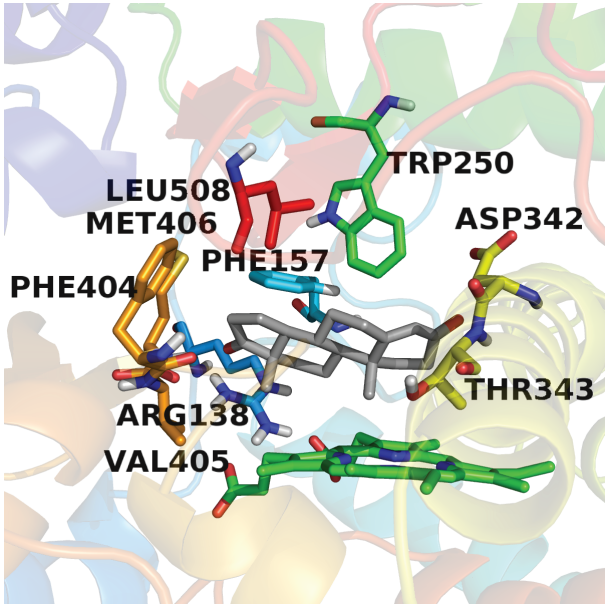

Human
-93 kb
-33 kb
0 kb
30 kb
I.1
I.F
PII/II
III
IV
V
VI
VII-IX
X
Amphioxus
0 kb
15 kb
Zebrafish A
(CYP19A1)
Zebrafish B
(CYP19A2)

```
ORGANISM                I-HELICAL DOMAIN                AROMATIC REGION          HEME BINDING DOMAIN
                . * *.: *:::*.***:*:.::.:*. : .   :*** : **:: .:* ::*  : :***:*:*  * *** * *.*.*:
Amphioxus(Bf)   SCVRQCVTEMLVAGPDTMSVNIYFILLHIAE   RPVVTFVMRHAEEEDHVDGYVIPKGTNVIINL  FMPFGLGVRSCVGRTIAP
Amphioxus(Bb)   ..............................   .....L...R..............A......  .....P..........
Dogfish         DN...SML.I.I.G.....S.F.M.ML..Q   Q...D.T..K.LKD.VI...PVK....I.L.L  .Q...C.P.S...KY..M
ZebrafishB      DD.....L..VI.A...L.ISLF.M..LLKQ   H...D.I..QSL.D.YI...RVA....L.L.I  .Q...C.P.A...KH..M
KillifishB      DD.....L..VI.A...L.ISLF.M.MLLKQ   H...D.T..RALDD.D.E.TK.K....I.L.I  .Q...C.P.S...KH..M
ZebrafishA      EN.....L..VI.A...L.ISLF.M..LLKQ   H...D.T..R.LDD.VIE..NVK....I.L.V  .Q...S.P.A.A.KY..M
KillifishA      EN.T...L..VI.A...L.ISLF.M..LLKQ   H...D.T..R.LSD.VI...RV.....I.L.T  .Q...S.P.A...KH..M
Xenopus         EN.N.CIL...I.A...M..SLF.M.VL..Q   Q...DL...K.L.D.II...YVK....I.L..  .Q...S.P.A.A.KY..M
Chicken         EN.N...L..MI.A...L..TLFIM.IL..D   Q...DLI..K.LQD.VI...PVK....I.L.I  .Q...F.P.G...KF..M
Zebrafinch      EN.N...L..MI.A...L..TLF.M.IL...   Q...DLI..K.LQD.VI...PVK....I.L.I  .Q...F.P.S...KF..M
Turtle          EN.N...L..MI.A...L..TLF.M.VL...   Q...DL...K.LQD.VI...PVKR...I.L.I  .Q...F.P.G...KF..M
Alligator       EN.N...L..MI.A...L..TLF.M.VL...   Q...DLI..K.LQD.VI...PVK....I.L.I  .Q...F.P.A...KF..M
Rat             EN.N..IL...I.A.....TL.VM..L...   Q...DL...R.L.D.VI...PVK....I.L.I  .Q...F.P.S.A.KY..M
Pig             EN.N..IL...I.A...L..TVF.M.FL..K   Q...DL...K.L.D.VI...PVK....I.L.I  .Q...F.P.A.A.KY..M
Human           EN.N..IL...I.A.....SLF.M.FL..K   Q...DL...K.L.D.VI...PVK....I.L.I  .Q...F.P.G.A.KY..M
```

Figure 3

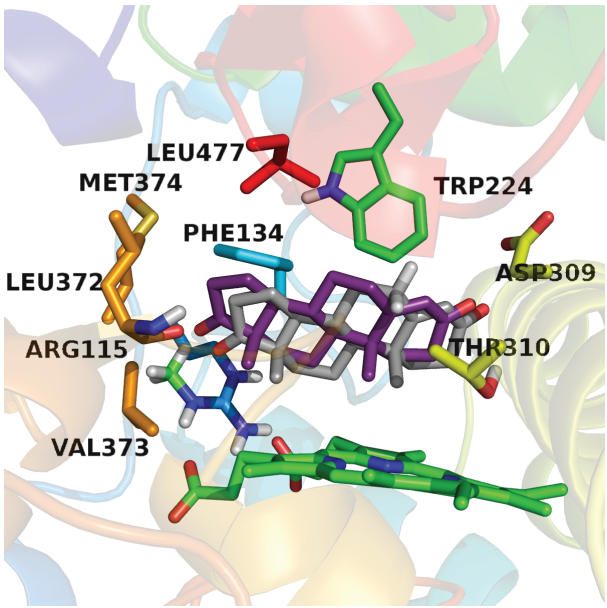
Amphioxus

A.

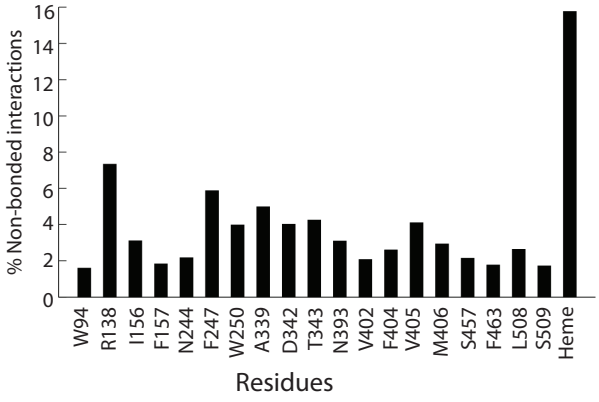


Human

B.



C.



D.

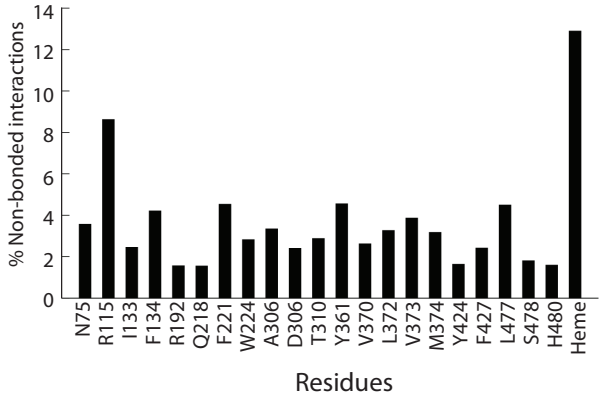
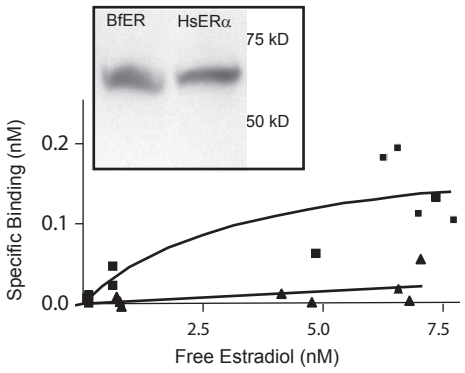


Figure 5

A.


BfER
HsERα
75 kD
50 kD
Specific Binding (nM)
0.2
0.1
0.0
2.5
5.0
7.5
Free Estradiol (nM)

B.

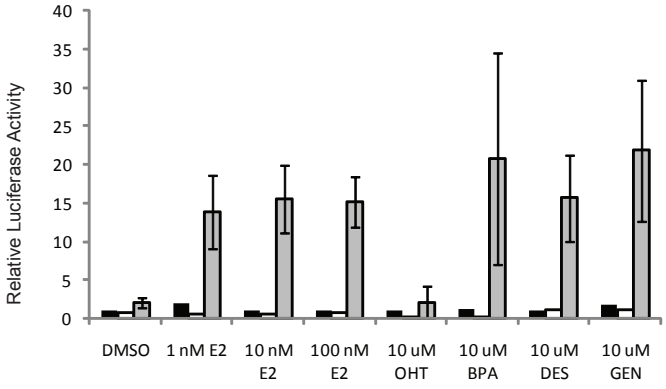
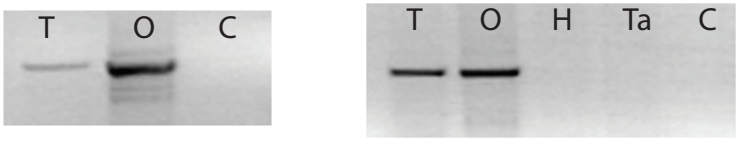

Relative Luciferase Activity
40
35
30
25
20
15
10
5
0
DMSO
1 nM E2
10 nM E2
100 nM E2
10 uM OHT
10 uM BPA
10 uM DES
10 uM GEN

Figure 6

A.



B.

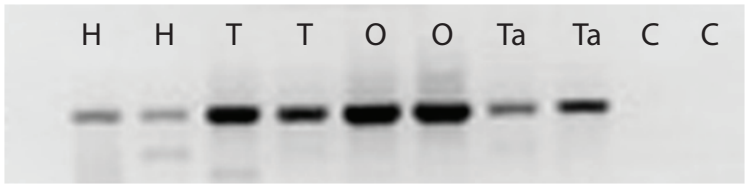


Table 1. Comparison of exon and intron sizes in coding region of amphioxus and human *CYP19* genes.

	Size (bp)	
Exon	amphioxus	Human
II	219	183
III	151	151
IV	155	155
V	186	177
VI	118	115
VII	133	115
VIII	160	163
IX	245	242
X	1,413	1,582
Intron		
II	332	5,758
III	368	8,925
IV	1,516	5,253
V	292	3,693
VI	338	2,723
VII	379	470
VIII	459	2,508
IX	463	1,263

Table 2. **Position and sequence of putative transcription factor binding motifs identified in the 5'-flanking region of the amphioxus *CYP19* gene.** Putative cis regulatory elements were identified using the MATCH program. Core and matrix match values indicate the quality of a match between the sequence and the cis regulatory element matrix, with 1.0 being a complete match [69]; where multiple matrices match a given site, the highest quality match is indicated.

Putative TF Match Quality (core/matrix)	Position from ATG (+/-) Strand	Sequence	Identification in other *CYP19* promoters [reference]
TATA 1.0/0.997	-187 to -195 (+)	TATAAAAA	goldfish and zebrafish A1 and A2 [19,89]; human I.3, 1.6 [99,100], zebra finch 1b [101], mouse [102,103]
ERR; ER ½ 1.0/0.918	-319 to -348 (+)	ATGTGTCTTTT(TGAC C)TCTGCATATTACT	EREs in goldfish and zebrafish A1, A2 [19,104]; mouse II and human 1f [reviewed in 104]
OCT-1 1.0/0.969	-319 to -331 (-)	TCTGCATATTACT	zebra finch 1a [101]
CREB 1.0/0.985 1.0/0.948	-354 to –363 (+) -637 to -642 (-)	TGACGTCT ACACGTCATAGG	zebrafish A1 and A2 [89]; rat II ([105]), Human I.3/II [106]
ERR; SF-1*	-461 to -469 (-)	TCAAGGTTA	goldfish and zebrafish A1 [19,104], human II [107], zebra finch 1b [101]
GATA1/GATA2 1.0/0.987	-430 to -444 (+)	AACAAAGATAAGTGT	zebra finch 1b [101], goldfish A2 [19], human endothelial 1.7 [28]

* Identified through a manual search; nearly perfect match with consensus sequence (TCAAGGTCA [95])

Supplementary Table 1. Oligonucleotide primers. The nucleotide position of sequences targeted by primers A1-30 and E1-6 refers to newly isolated amphioxus aromatase and ER cDNA sequences (GenBank Accession Numbers: DQ165086 and EF554313, respectively) or to sequence obtained in whole or in part by interrogation of the amphioxus genome database (*).

Primer	Application	Sequence 5' – 3' (F, forward; R, reverse)	(nt)
A1	cDNA cloning	F-ATTGAACAGAAGTCGTACCAGACA	47-70
A2	cDNA cloning	R-GGTGTTTCCTGTACAGGGTAGGAT	796-819
A3	cDNA cloning	R-AAGACTTCCCTGATTTCCCTAAGT	1097-1120
A4	cDNA cloning	F-TACCAGACACGGCATAACCA	62-81
A5	cDNA cloning	R-GCAGTAGCGTGATGAGGACA	1450-1469
A6	3' RACE	F-GCAAGGCCAGTGGTCACTTTCGTCAT	1197-1222
A7	5' RACE	R-AACTTCCAGCTCTGCACGTCGTTGTT	476-502
A8	5' RACE	R-CGGATGTTAGCCAGTTGTCGTCTGGT	576-601
A9	RT-PCR	F-AGGACAGGAGTCAATTATCA	30 – 49*
A10	RT-PCR	R-CGCTACTAGGACAGAAGAAG	1981-1999
A11	RT-PCR	F-AAAGCAGAGGTAGCTTCCCATT	28-39
A12	RT-PCR	R- TGGATACTTTCCCTCAGATCGT	1581-1602
A13	Intron II	F-GGGATATTTCCCCCAGGTAA	204-221
A14	"	R-CTGCTGATGATGAAAGTCTGCT	349-370
A15	Intron III	F-CATCAGCAGGTGCGTTCTTA	362-386
A16	"	R-GCACGTCGTTGTTGAAGATG	470-489
A17	Intron IV	F-CGCTTCTTCTTCGTCAAAGG	507-525
A18	"	R-AAATAGCCGGTTTCCGATTT	670-688
A19	Intron V	F-ATCGGAAACCGGCTATTTCT	672-691
A20	"	R-GATAAACCAACTGCAGCACCT	778-798
A21	Intron VI	F-GGAGGTGGTGAACACAATCA	716-735
A22	"	R-TTGGGATGCTGAATTCTTCC	902-921
A23	Intron VII	F-GGAAGCTGGTGGACAAGAAG	853-872
A24	"	R-ACATTTCTGTCACGCACTG	996-1012
A25	Intron VIII	F-CAGTGCGTGACAGAAATGCT	996-1015

A26	“	R-ATCGAGGAACACCATCTTGC	1153-1172
A27	Intron IX	F-CACCAAGACCCTCGTCACTT	1305-1324
A28	“	R-TCCCTCAGATCGTCTCCAAC	1573-1593
A29	Promoter	F-ACGAACACAACCAATCCAG	1164 – 1184*
A30	“	R-ACAAGCACAACAAGCAGCAC	134-156
E1	cDNA cloning	F-TGGAGTCTGGTCATGTGAGG	968-987
E2	cDNA PCR	R-CAGTTTCTCCCCCTTGTGTC	1642-1661
E3	5’ RACE	R-GGCAGCTTTTTCGTCTATTCCTGTCAATG	1064-1092
E4	3’ RACE	F-TGGGGACCATGCTTCAGGCTTCCAC	941-965
E5	RT-PCR	F-CATGTCTCCTGTCATGTCTGTC	656-673
E6	RT-PCR	R-GCAATCATCTCTCTTTCTCTGG	2142-2163

*bp upstream of ATG translation start site