# Gene structure of the novel cytochrome P4501D1 genes in stickleback (*Gasterosteus aculeatus*) and medaka (*Oryzias latipes*)

**Jared V. Goldstone** and **John J. Stegeman**[*]
Biology Department, Redfield 3-42 MS#32, Woods Hole Oceanographic Institution, Woods Hole MA, 02543 USA

## Abstract

The cytochrome P450 1 (*CYP1*) family has expanded with the addition of the *CYP1B* and *CYP1C* subfamilies. We recently identified a new CYP1 subfamily in zebrafish, *CYP1D*, with a single gene, *CYP1D1*. Here we examined sequences found in other fish genomes, *i.e.*, stickleback (*Gasterosteus aculeatus*) and medaka (*Oryzias latipes*), for similarities among fish *CYP1D1* genes. The full-length deduced amino acid sequences for CPY1D1 in these two species averaged about 43% identity to the CYP1As, but nearly 50% when sequence alignment ambiguities were masked. *CYP1D1* has seven exons, similar in size and position to the exons in *CYP1D1* and *CYP1A* in zebrafish. However, the intronic distances were substantially smaller in the medaka and stickleback. There also were differing numbers of xenobiotic response elements in the *CYP1D1* of the various species. Whether the stickleback or medaka genes are inducible by aryl hydrocarbon receptor (AHR) agonists is yet to be determined.

## Keywords

Fish; isozymes; microsomal oxidases; CYP1A; bioinformatics; ortholog

Members of the CYP1 of monooxygenase family catalyze the oxidation and often bioactivation of a wide variety of common environmental carcinogens and promutagens. CYP1 substrates include aromatic hydrocarbons, especially polycyclic aromatic hydrocarbons such as benzo[a]pyrene, as well as aromatic amines and heme degradation products (Nebert et al., 2004). Planar PCB congeners such as 3,3′4,4′-tetrachlorobiphenyl are also substrates, although metabolized very slowly (White et al., 1997). In mammals, three *CYP1* genes have been described: *CYP1A1, CYP1A2* and *CYP1B1*. Homologous *CYP1A* and *CYP1B* genes also have been described in fish. In addition, however, fish also express two paralogous genes in the *CYP1C* subfamily, *CYP1C1* and *CYP1C2*, not found in mammals (Godard et al., 2005).

Recently we identified a fourth *CYP1* subfamily in fish, *CYP1D*, with a single gene *CYP1D1*. *CYP1D1* transcript is expressed in many organs of adult zebrafish but appears to be most highly expressed in early embryos. *CYP1D1* transcript is not induced by the potent

---

[*]Corresponding author: Tel: 508-289-2320; fax: 508-457-2341, jstegeman@whoi.edu (J.J. Stegeman).

AHR agonist PCB 126 (3,3′,4,4′,5-pentachlorobiphenyl), implying a possible endogenous role during development (Goldstone et al., submitted). The amino acid sequence of CYP1D1 in zebrafish inferred from cDNA and from genomic sequence is more like that of CYP1A than of CYP1B1 or the CYP1Cs. We showed that the differences in sequence between CYP1D1 and CYP1A are not evenly distributed along the protein, but that there are regions of higher or lower amino acid similarity, as expected. We also have found *CYP1D1* sequences in other fish species. Here, we compare the sequence identities and gene structure of *CYP1D1* in medaka (*Oryzias latipes*) and three-spined stickleback (*Gasterosteus aculeatus*), showing that *CYP1D1* resembles *CYP1A* also in these species.

The sequences for stickleback and medaka genomes and gene predictions were obtained from the Ensembl database (Hubbard et al., 2007) and searched using BLAST. Gene predictions for *CYP1A*, *CYP1B*, *CYP1C1*, *CYP1C2*, and *CYP1D1* were refined using Genscan and FGeneSH. Exon mapping was performed using GCG (v. 10.3; Accelrys, San Diego, CA).

The percent identity of amino acid sequences inferred for *CYP1D1* and *CYP1A* genes in stickleback, medaka and zebrafish are shown in Table 1. In these species the unambiguously aligned CYP1D1 sequences are 47 to 50% identical to the CYP1As. In comparison, fish CYP1B and CYP1C sequences average only 43 and 41% amino acid identity to CYP1As, respectively. Similar to zebrafish, medaka and stickleback CYP1D1 deduced amino acid sequences display regions of higher and lower sequence similarity to CYP1A (data not shown). Some of these regions of high similarity correspond to structural elements common to many CYPs, including the A, B, E, G, I, K, and L α-helices and β-sheets 3 and 4.

Regions showing higher sequence similarity between CYP1D1 and CYP1A include putative substrate recognition sites (SRS; Gotoh, 1992) 4, 5 and 6. In contrast, the sequences of SRS 2 and 3 are quite different between CYP1D1 and CYP1A, and exhibit lower sequence similarity than immediately surrounding regions of the protein (Goldstone et al., submitted). The differences in these two SRSs imply differences in substrate preferences between CYP1A and CYP1D1, assuming CYP1D1 is translated into protein.

Examination of the *CYP1D1* gene structure revealed seven exons in both medaka and stickleback (Figure 1), with a relative arrangement highly similar to the exon-intron structure of the *CYP1As* in these species, quite like the similarity detected in zebrafish. Furthermore, the exon-intron boundaries are the same in the *CYP1D1s* and the *CYP1As*, with the exception of the first part of the coding region in *CYP1D1* being split over the first two exons, while the first exon in *CYP1A* is non-coding. The *CYP1D1* genes in medaka and stickleback are compressed on the chromosome relative to the zebrafish *CYP1D1* (Figure 1). This may be a consequence of the smaller genome sizes in medaka (816 Mb) and stickleback (714 Mb) relative to zebrafish (1.7 Gb) (Ishikawa, 2000).

Mapping of consensus AHR binding sites in the proximal promoter regions (i.e., xenobiotic response elements or XREs; KNGCGTG), shows that medaka *CYP1D1* has zero and stickleback *CYP1D1* has four putative XREs in the upstream regulatory region (Figure 1); zebrafish *CYP1D1* has two putative XREs (Goldstone et al., submitted). In contrast, the AHR2-regulated zebrafish genes *CYP1A*, *CYP1B1*, *CYP1C1*, and *CYP1C2* have 12, 6, 4, and 6 XREs, respectively (Jönsson et al., 2007a, 2007b). Whether XREs in *CYP1D1* are functional is not known; not all XREs in zebrafish *CYP1A* are functional (ZeRuth & Pollenz, 2005, 2007).

The *CYP1D* and *CYP1A* subfamilies apparently diverged from one another subsequent to the divergence leading to the *CYP1B* and *CYP1C* subfamilies. Thus, within the vertebrate *CYP1* gene family, the *CYP1As* and *CYP1Ds* together constitute one clade, while the

*CYP1Bs* and *CYP1Cs* constitute another. As we have noted (Goldstone et al., in press, submitted), establishing the timing of the divergences leading to the four subfamilies will require further taxonomic sampling.

The similarity of *CYP1D1* to *CYP1A* has intriguing functional implications. However, at this time the expression of a CYP1D1 protein and any functions thereof has yet to be established. If the relatively greater levels of *CYP1D1* transcript expression in zebrafish embryos are accompanied by protein expression, it would suggest that there may be endogenous functions accomplished by CYP1D1, although CYP1D1 could have catalytic functions involving the transformation of xenobiotics. Interestingly, *CYP1D1* has not been found in either of the two pufferfish genomes [*Takifugu rubripes* or *Tetraodon nigroviridis*; (Nelson, 2003; Goldstone et al., submitted)]. Clearly, determining expression and catalytic functions of CYP1D1 protein is an important objective; such studies are underway.

## Acknowledgments

## References

Godard CA, Goldstone JV, Said MR, Dickerson RL, Woodin BR, Stegeman JJ. Biochemical and Biophysical Research Communications. 2005; 331:1016–1024. [PubMed: 15882979]

Goldstone JV, Goldstone HM, Morrison AM, Tarrant, Kern ASE, Woodin BR, Stegeman JJ. Molecular Biology and Evolution. 200710.1093/molbev/msm200

Goldstone JV, Jönsson ME, Woodin BR, Nelson DR, Stegeman JJ. Gene. (submitted).

Gotoh O. Journal of Biological Chemistry. 1992; 267:83–90. [PubMed: 1730627]

Hubbard TJ, Aken BL, Beal K, Ballester B, Caccamo M, et al. Nucleic Acids Research. 2007; 35:D610–D617. [PubMed: 17148474]

Ishikawa Y. Bioessays. 2000; 22:487–495. [PubMed: 10797489]

Jönsson ME, Jenny MJ, Woodin BR, Hahn ME, Stegeman JJ. Toxicological Science. 2007a; 100:180–193.

Jönsson ME, Orrego R, Woodin BR, Goldstone JV, Stegeman JJ. Toxicology and Applied Pharmacology. 2007b; 221:29–41.

Nebert DW, Dalton TP, Okey AB, Gonzalez FJ. Journal of Biological Chemistry. 2004; 279:23847–23850. [PubMed: 15028720]

Nelson DR. Archives of Biochemistry and Biophysics. 2003; 409:18–24. [PubMed: 12464240]

White RD, Shea D, Stegeman JJ. Drug Metabolism and Disposition. 1997; 25:564–572. [PubMed: 9152595]

ZeRuth G, Pollenz RS. Zebrafish. 2005; 2:197–210. [PubMed: 18248194]

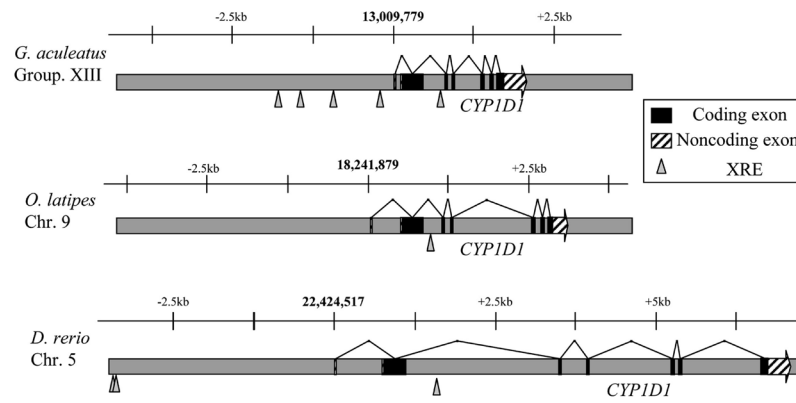ZeRuth G, Pollenz RS. Chemico-Biological Interactions. 2007; 170:100–113. [PubMed: 17720151]

**Fig. 1.**
Comparison of the gene structure for fish *CYP1D1*. Expressed exons are black boxes, while striped boxes indicate untranslated regions. Shown also are the locations of calculated XRE sequences (triangles; sequence searched is KNGCGTG). Bold numbers refer to the location of the calculated transcriptional start site on the genome assembly. Scale bars are equivalent for all three genes depicted.

**Table 1**

Percent predicted amino acid identity between zebrafish (DANRE), stickleback (GASAC) and medaka (ORYLA) CYP1D1 and CYP1A sequences (lower left) and with regions of ambiguous alignment excluded by masking (upper right).

| | DANRE_1D | GASAC_1D | ORYLA_1D | DANIO_1A | ORYLA_1A | GASAC_1A |
|---|---|---|---|---|---|---|
| DANRE_1D | | 76.6% | 73.3% | 49.0% | 48.3% | 48.1% |
| GASAC_1D | 67.1% | | 83.7% | 47.7% | 48.3% | 47.0% |
| ORYLA_1D | 63.6% | 74.1% | | 49.7% | 50.2% | 47.4% |
| DANRE_1A | 44.4% | 40.8% | 42.0% | | 73.3% | 75.4% |
| ORYLA_1A | 42.9% | 42.2% | 42.3% | 69.6% | | 77.7% |
| GASAC_1A | 42.7% | 40.3% | 40.4% | 72.5% | 75.2% | |