ARCHIVES OF BIOCHEMISTRY AND BIOPHYSICS
Vol. 322, No. 1, September 10, pp. 204–213, 1995

Characterization and Prevalence of a Polymorphism in the 3′ Untranslated Region of Cytochrome P4501A1 in Cancer-Prone Atlantic Tomcod

Nirmal K. Roy, Guat-Lian Kreamer,[1] Blake Konkle, Cheryl Grunwald, and Isaac Wirgin[2]
Nelson Institute of Environmental Medicine, New York University Medical Center, Long Meadow Road, Tuxedo, New York 10987

Received April 28, 1995, and in revised form June 14, 1995

Atlantic tomcod (*Microgadus tomcod*) from the cancer-prone Hudson River population exhibit a genetic polymorphism in the cytochrome P4501A (CYP1A) gene which is evidenced in Northern blot analyses by a truncated transcript and in Southern blot analyses by a deletion in the variant allele. To initially evaluate the functional significance of this polymorphism, we sought to characterize the molecular basis for this polymorphism and to determine its frequency in tomcod from other populations in which the prevalence of neoplasia is low. The common CYP1A allelic sequence was determined from β-naphthoflavone-induced tomcod cDNA and from tomcod genomic DNA. A sequence of the variant CYP1A allele was obtained by direct sequence analysis of the amplicons of variant tomcod cDNA and genomic DNA. CYP1A exon and intron structure is highly conserved between tomcod and all other teleost and mammalian species compared. Similarity of the deduced tomcod, rainbow trout, and plaice amino acid sequences was 72%, whereas similarity between tomcod and mammalian sequences was approximately 50%. The variant tomcod CYP1A allele results from a 606-bp deletion in the 7th exon of the 3′ untranslated region (UTR) of the cDNA. Polymerase chain reaction and Northern blot analyses revealed an absence of this CYP1A polymorphism in tomcod from other rivers. Studies in humans suggest that variation in CYP1A1 cDNA may impact on genetic susceptibility to environmentally induced neoplasia. Furthermore, studies in *in vitro* mammalian models indicate the importance of 3′ UTRs on gene expression by impacting on the stability of transcript. These results suggest that the 3′ UTR CYP1A polymorphism in tomcod may have consequences for the genetic susceptibility of Hudson River fish to hepatic neoplasia.

Key Words: **CYP1A1; 3′ UTR; Atlantic tomcod; genetic polymorphism; genetic susceptibility.**

Both environmental contamination and genetic susceptibility contribute to the prevalence of chemically induced neoplasms in natural populations (1). Associations between increased prevalences of hepatic lesions in feral fish populations and elevated sediment concentrations of xenobiotics such as polynuclear aromatic hydrocarbons (PAHs)[3] have recently been reported (2, 3). The relationship between the prevalence of liver lesions and sediment concentrations of halogenated aromatic hydrocarbons (HAHs) such as PCBs is not as clear (4), although these compounds have also been implicated as risk factors in some fish studies (5).

Atlantic tomcod (*Microgadus tomcod*) from the Hudson River, New York estuary are exposed to unusually high concentrations of both classes of aromatic hydrocarbon compounds and exhibit an unusually high prevalence of liver tumors. Approximately 50% of 1-year-old and more than 90% of 2-year-old tomcod display hepatocellular carcinomas (6), in comparison to a prevalence of 5 and 0%, respectively, in tomcod from cleaner rivers in Maine (7) and the Canadian Maritime provinces (S. Courtenay, unpublished data). The Hudson

[1] Present address: DuPont CR & D, Haskell Laboratory, P.O. Box 50, Newark, DE 19714.

[2] To whom correspondences should be addressed. Fax: (914) 351-5472. E-mail: wirgin@charlotte.med.nyu.edu.

[3] Abbreviations used: CYP1A1, cytochrome P4501A1; PAH, polynuclear aromatic hydrocarbon; PCB, polychlorinated biphenyl; dioxin, 2,3,7,8-tetrachlorodibenzo-*p*-dioxin; B[*a*]P, benzo[*a*]pyrene; HAHs, halogenated aromatic hydrocarbons; 3-MC, 3-methylcholanthrene; RT-PCR, reverse transcriptase polymerase chain reaction; SSC, standard sodium citrate; SDS, sodium dodecyl sulfate; nt, nucleotide; UTR, untranslated region; β-NF, β-naphthoflavone.

0003-9861/95 $12.00

River has a long history of environmental insult from both point and nonpoint sources of xenobiotic discharge (8). As a result, sediments from the Hudson River contain among the highest levels of PAHs, PCBs, and dioxins in any U.S. estuary (9). Elevated hepatic tissue concentrations of these compounds have been detected in several Hudson River fish species (10), including high levels of PCBs in tomcod (11). Exposure of Hudson River tomcod to high levels of PAHs has been confirmed by detection of PAH metabolites in bile and by high levels of expression of hepatic cytochrome P4501A (CYP1A) mRNA (12, 13). Furthermore, elevated levels of hepatic DNA adducts detected by ^{32}P postlabeling analysis (13) and an activated K-*ras* oncogene in Hudson River tomcod liver tumor DNA (14) suggest that exposure to PAHs has resulted in hepatic DNA damage at carcinogenically relevant gene loci. However, no other species of Hudson River fish examined exhibit hepatic neoplasms and inducibility of CYP1A mRNA differs significantly among finfish species (15). Furthermore, not all tomcod in the Hudson River exhibit liver neoplasms and variation among individuals in CYP1A mRNA inducibility is high (16).

CYP1A gene expression in fish is induced by exposure to PAHs, coplanar PCB congeners, dibenzo-*p*-dioxins, and dibenzofurans (17). Cytochrome P4501A1-encoded monooxygenases transform environmental procarcinogenic PAHs to reactive metabolites. These metabolites can conjugate with carrier molecules which expedite their elimination from the body or alternatively react with cellular macromolecules such as proteins and DNA. Formation of DNA adducts is believed to be an early and necessary step in the initiation of chemically induced multistage carcinogenesis (18). Induction of CYP1A gene expression by HAHs may also play a significant role in the promotion and progression of neoplasia in previously initiated hepatic cells (19).

We have previously reported that levels of CYP1A mRNA are highly induced in tomcod from contaminated sites (12). For example, levels of CYP1A mRNA expression were 28-fold higher in tomcod from the Hudson River compared to tomcod collected from the pristine Margaree River, Nova Scotia (13). CYP1A mRNA levels were also elevated (up to 5.5-fold) in caged tomcod exposed to bleached kraft mill effluent in the Miramichi River, New Brunswick (20). During the course of these studies, we observed a genetic polymorphism in CYP1A transcript in tomcod from the Hudson River that was absent in a small sample of tomcod from two relatively pristine rivers in Maine (21). Initially, we detected a variant truncated CYP1A mRNA by Northern blot analysis in approximately 10% of Hudson River tomcod and the genetic basis of this polymorphism was confirmed by RFLP analyses with a battery of restriction endonucleases.

In this study, we sought to determine the molecular basis of this polymorphism and the frequency of variant CYP1A genotypes in tomcod from a larger number of rivers encompassing most of the species' distribution. We provide the sequence of the common tomcod CYP1A allele, its exon–intron structure, its similarity to CYP1A alleles reported in other species of fish and mammals, and a comparison of the common and variant tomcod CYP1A alleles. This study represents the first evidence for a genetic polymorphism in CYP1A in a teleost species which varies widely in the prevalence of hepatic neoplasia among natural populations.

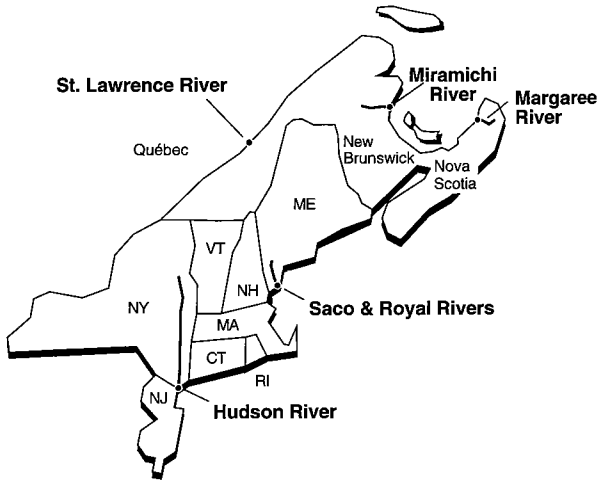


FIG. 1. A map depicting North American Atlantic coast rivers from which Atlantic tomcod were collected for this study.

MATERIALS AND METHODS

Sample collections. Adult tomcod were collected from the six Atlantic coast rivers depicted in Fig. 1. Of these rivers, the Hudson River (n = 268) is highly contaminated with PCBs, dioxins, furans, and PAHs; the Saco and Royal Rivers in Maine (n = 24) are relatively pristine, the St. Lawrence River (n = 48) has high levels of PAHs and possibly other classes of aromatic hydrocarbon compounds, the Miramichi River (n = 44) receives localized discharge from a bleached kraft pulp and paper mill, and the Margaree River (n = 45) has no industrialization or urbanization.

DNA analysis. DNA was isolated from tomcod livers which were homogenized and incubated in NH_4OH/Triton X-100 solution as described by Wirgin *et al.* (22). DNA concentrations and purity were determined spectrophotometrically. DNAs from all samples were digested with *Bam*HI as previously described (21) to determine CYP1A genotypes. Digested DNAs were electrophoretically separated in 1.0% agarose gels, transferred to ZetaBind membranes (Cuno, CT), and analyzed in Southern blot analyses (23). Membranes were hybridized at 65°C (24) to a ^{32}P-radiolabeled (25) plasmid containing 1.4 kb of 3-methylcholanthrene-induced rainbow trout CYP1A1 cDNA (26). Final wash conditions were 1.0 × SSC/0.1% SDS at 65°C for 1 h. CYP1A hybridizable DNA fragments were visualized by autoradiography.

In subsequent years, population screening for CYP1A variants was accomplished by PCR amplification of 500 ng of genomic DNA isolated from 1-cm^2 pieces of pectoral fin tissue as previously described (15). A 40-μl reaction was cycled as follows: initial 4 min denaturation

at 95°C; 25 cycles of 1 min denaturation at 94°C, primer annealing at 52°C, and extension at 72°C; additional final extension for 7 min at 72°C. Using primers P3 and P4 (Table I), common and variant genotypes were differentiated on ethidium bromide-stained 1.5% agarose gels.

RNA isolation and analysis. Total RNA was prepared from livers of tomcod treated (ip injected) with 10 mg/kg fish benzo[*a*]pyrene (B[*a*]P) or β-naphthoflavone (β-NF); 100 μg/kg fish 3,3′,4,4′-tetrachlorobiphenyl; or 50 ng/kg fish 2,3,7,8 tetrachlorodibenzo-*p*-dioxin using the method of Chomczynski and Sacchi (27) and RNAzol reagent (Biotecx Laboratories, TX). RNAs were denatured with formaldehyde and analyzed by electrophoresis in 1.0% agarose gels. Gels were stained with ethidium bromide to evaluate the integrity of rRNA in each lane. Northern blot hybridizations were conducted exactly as described (13) and final wash conditions were 1.0 × SSPE/0.1% SDS at 65°C for 1 h.

A cDNA library was developed from a β-NF-treated tomcod which was originally collected from the Hudson River. Poly(A)$^+$ RNA was purified from approximately 500 μg of total RNA using oligo(dT) cellulose columns (Life Technologies, MD). First-strand cDNA was synthesized from approximately 5 μg of poly(A)$^+$ RNA using Moloney murine leukemia virus reverse transcriptase and oligo(dT)$_{12-18}$ according to the manufacturer's instructions (Life Technologies). Second-strand synthesis was carried out as described (28), products were purified and size-fractionated on CL-4B columns (Life Technologies) to select for fragments larger than 500 bp, and these were ligated to an *Eco*RI (*Not*I) adaptor (Life Technologies) and inserted in the *Eco*RI site of λgt-10 (Stratagene, CA). Ligation products were packaged using Gigapack II extract (Stratagene). The library contained 1.2×10^6 independent recombinant clones that were subsequently amplified and stored at 4°C. The library was screened by plaque hybridization with the ^{32}P-radiolabeled (29) rainbow trout CYP1A cDNA probe (26).

Development of tomcod genomic DNA library. A tomcod genomic DNA library was developed from partially digested (*Sau*3A) and size-fractionated DNA isolated from the liver of a tomcod collected from the Hudson River. DNA fragments (9–20 kb) purified from a continuous sucrose gradient (28) were ligated to *Bam*HI-digested λ-Dash II arms (Stratagene). Recombinant DNA was packaged with Gigapack II extract and infected into XL1-Blue MRA (P2) host cells (Stratagene). The library was plated, titered, and amplified (28). Plaques were screened with the full-length ^{32}P-radiolabeled (29) tomcod CYP1A cDNA probe.

Hybridizations were overnight at 65°C and final washes were 0.1 × SSC/0.1% SDS at 65°C for 1 h. One CYP1A hybridizable recombinant was identified in 10^6 plaques and it was purified through tertiary screening. *Bam*HI was used to isolate the tomcod genomic DNA insert from vector and *Pst*I was used to generate DNA fragments that were subcloned into pUC19 for dideoxy sequencing using Sequenase Version 2.0 (U.S. Biochemicals, OH) and [^{35}S]dATP. Selection and orientation of these genomic DNA fragments were determined by hybridization to the tomcod CYP1A cDNA probe and restriction mapping. Additional subcloning and primer walking with the oligonucleotides listed in Table I were used to obtain the complete tomcod CYP1A genomic DNA sequence.

Sequence and statistical analyses. Sequence analyses were performed using GCG software (Genetics Computer Group, Madison, WI). Initially, similarities between the sequences of tomcod cDNA subclones and CYP1A1 sequences in GenBank were determined and multiple sequences were then aligned to obtain a contiguous consensus sequence. The Translate program was used to deduce the amino acid sequence and this was compared to published CYP1A1 and CYP1A2 sequences. The Mfold program was used to determine secondary structure of the 3′ untranslated region of the cDNA. This method uses the energy rules developed by Turner *et al.* (30) to determine optimal secondary structures.

Frequencies of CYP1A1 genotypes were determined for tomcod populations and differences were tested for statistical significance using the Monte Carlo-based χ^2 method of Roff and Bentzen (31) which permits robust analysis of data sets including small cell sizes.

Primer extension analysis. Tomcod CYP1A1 primer P12 was end-labeled with [γ-^{32}P]ATP using T4 polynucleotide kinase (New England Biolabs, MA). Labeled primer (0.5 pmol) was added to 5 μg of tomcod total hepatic RNA in 10 μl of 10 mM Tris–Cl buffer (pH 7.6) containing 1 mM $MgCl_2$ and 0.5 mM dithiothreitol, heated at 90°C for 2 min, and cooled slowly to 45°C. Samples were freeze-dried and resuspended in 10 μl of reaction mix containing 100 mM Tris–Cl (pH 8.3), 40 mM KCl, 10 mM $MgCl_2$, 0.4 mM each of dATP, dGTP, dCTP, and dTTP, 5 units of RNase inhibitor (Invitrogen, CA), 20 units of RNase Block I (Stratagene), and 5 units of AMV reverse transcriptase (Invitrogen). Reactions were incubated at 42°C for 45 min and terminated by the addition of 5 μl of 95% formamide, 20 mM EDTA, 0.05% bromophenol blue, and 0.05% xylene cyanol. Reactants were denatured at 95°C for 5 min and electrophoresed in denaturing 6% polyacrylamide gels. Lengths of the extension products were determined by comparison to a ladder of a recombinant tomcod CYP1A1 clone sequenced using primer P12.

RESULTS

CYP1A mRNA induction in Atlantic tomcod is highly sensitive to environmentally relevant concentrations of xenobiotics and by exposure to model CYP1A1 inducers

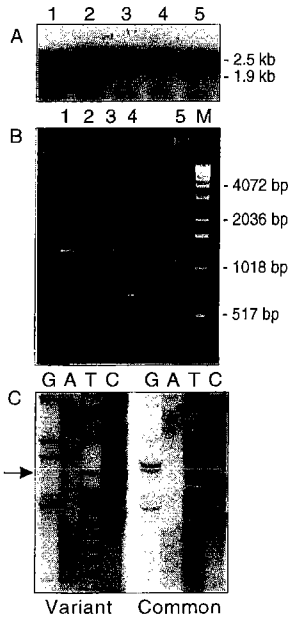


FIG. 2. (A) Northern blot analysis of CYP1A mRNA in tomcod injected with 10 ppm B[*a*]P and sacrificed after 48 h. Samples in lanes 1 and 2 exhibit the variant CYP1A mRNA, whereas those in lanes 3–5 display the common transcript. (B) An agarose gel with PCR products of tomcod genomic DNA and cDNA amplified with primers P3 and P4. Lanes 1 and 2 exhibit amplification of genomic DNA from fish with the common allele, lanes 3 and 4 show products of amplification of genomic DNA from tomcod with the variant genotype, lane 5 shows the PCR product of cDNA from a fish with the common genotype, and lane M contains a 1-kb molecular weight ladder. (C) A comparison of the sequence of the common and variant CYP1A alleles showing the site of the deletion at nt 1709 (arrow). Primer P4 was used for sequencing.

TABLE I

Tomcod-Specific Oligonucleotide Primers Used in PCR Amplification of Tomcod CYP1A cDNA and Genomic DNA

Primers	5′ to 3′ primer sequence	5′ Nucleotide position
P1	GTTGCCATGATCACCGTGTG	162
P2	GCTTCCTGTCCTCGCAGTGG	990
P3	ACTGTGCCACAAAAGACACG	1273
P4	AAAATACAAGGAAAGCAATA	2414
P6	GTAATGCTCTTCGGCATGGG	1452
P8	TGGAGTCAGTCTTCTTGA	1720
P9	TTTTTATTAATTTGTCAAAT	2347
P10	CCACCACCTTCCCGAA	795
P11	GGAAGCAGATATTAGGC	2109
P12	CACAACAAACGAGATCCC	79

such as the coplanar PCB congener 3,3′,4,4′-tetrachlorobiphenyl (32); 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (20); and the PAH benzo[*a*]pyrene (Fig. 2A, data not shown for the PCB congener and TCDD).

To isolate and characterize the CYP1A gene in Atlantic tomcod we screened a λgt-10 cDNA library prepared with poly(A)$^+$ mRNA isolated from a β-NF-exposed fish. Approximately, 10^6 phage were screened with the rainbow trout 3-methylcholanthrene-induced CYP1A1 cDNA probe. After tertiary screening, three phage clones, containing 2.5-, 2.4-, and 1.5-kb inserts, were found to be strongly positive in plaque hybridizations. All three fragments were subcloned into pUC19 and DNA sequences were determined using M13 forward and reverse primers and by primer walking with 10 tomcod-specific primers based on derived sequences (Table I).

All three clones contained a poly(A) tail but were extended to different lengths during first-strand synthesis. Sequence of the 2.5-kb tomcod cDNA fragment is presented in Fig. 3. This cDNA sequence includes a putative poly(A) signal sequence (AAATAAAA) which is 15 nucleotides (nt) upstream of the 14-nt-long poly(A) sequence (not shown). Comparison of the tomcod to CYP1A1 cDNA sequences in other fish species showed similarities of 78.1% with plaice (33) and 78.6% with rainbow trout (34). As expected, comparisons with mammalian CYP1A1 sequences showed lower levels of similarity. For example, the tomcod CYP1A cDNA sequence exhibited 66.1 and 65.4% similarity with human (35) and mouse (36) CYP1A1 cDNA sequences, respectively.

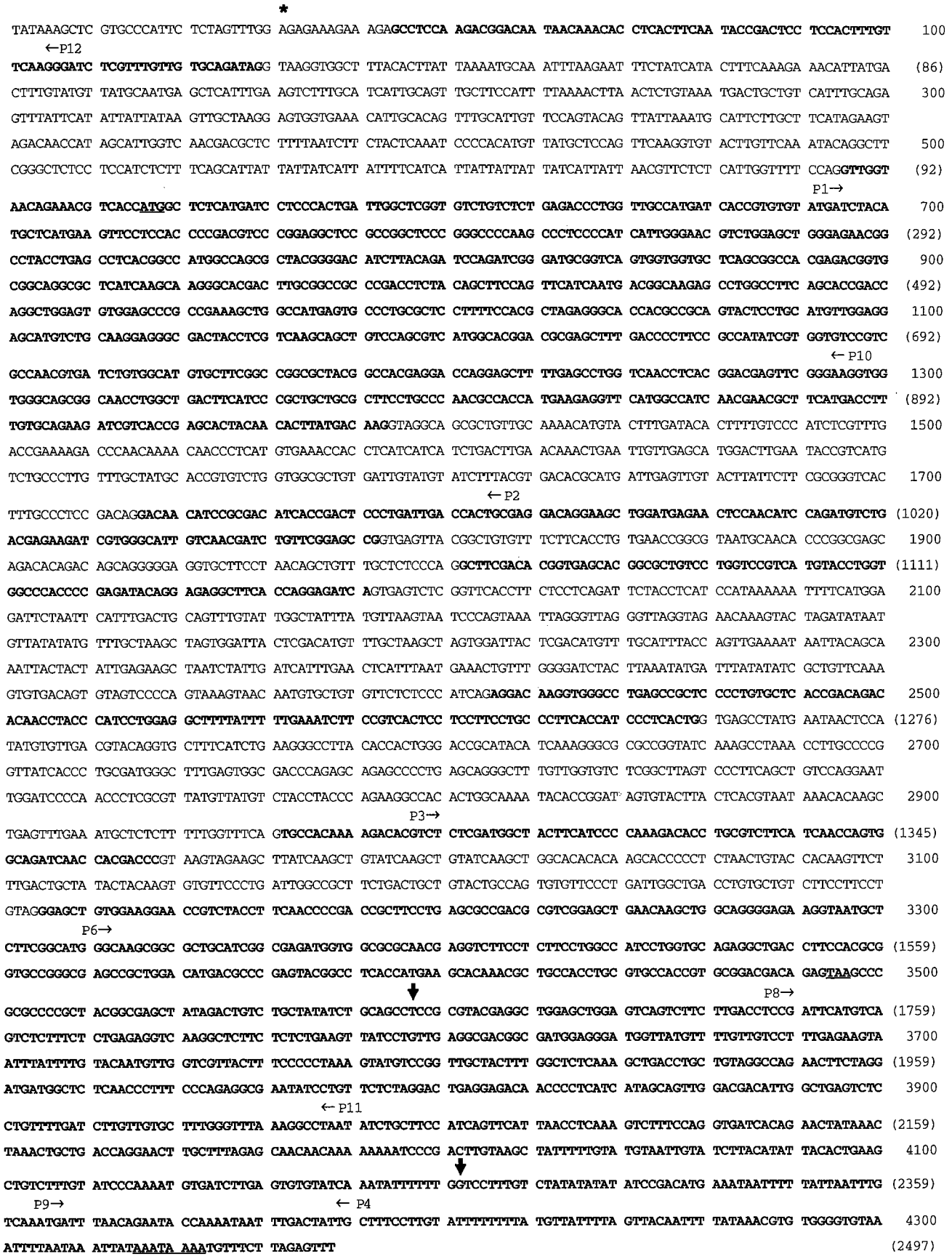
Tomcod CYP1A cDNA contains an open reading frame from nt 108 to nt 1655, which encodes for a peptide of 515 residues with a molecular weight of 57.8 kDa (Fig. 4). Comparison of the tomcod CYP1A amino acid sequence with that of plaice and rainbow trout in each case revealed about 72% identity. The tomcod amino acid sequence exhibited less identity with mammalian CYP1A1 sequences. For example, the highest level of identity was observed between tomcod and human (52%), and the lowest level of identity was between tomcod and hamster (49%). The putative tomcod CYP1A amino acid sequence showed less similarity to mammalian CYP1A2 sequences. Amino acid sequence similarity between tomcod CYP1A and mammalian CYP1A2 ranged between 45% (rat) and 49% (human).

To confirm the tomcod cDNA sequence and to characterize genomic DNA arrangements, a genomic DNA library was prepared in λ-Dash II and screened with the full-length 2.5-kb tomcod cDNA probe. Phage clones (10^6) were screened and a single positive plaque was identified and isolated, and its 12-kb insert was purified. Three *Pst*I restriction fragments (4.8, 1.5, and 3.6 kb) were identified which contained all CYP1A exons. Orientation of these fragments was determined by restriction mapping and Southern blot hybridization. Contiguous nucleotides (4338) were sequenced from these three genomic DNA fragments and these were found to contain all the exonic sequences identified from the tomcod CYP1A cDNA (Fig. 3). Upstream sequences (3 kb) containing putative CYP1A regulatory elements were characterized from the 4.8-kb fragment and will be described elsewhere (Roy *et al.,* manuscript in preparation). No differences were found between the tomcod CYP1A cDNA and exons in the tomcod CYP1A hybridizable genomic DNA.

Comparison of tomcod CYP1A1 intronic sequences with those of rainbow trout (34) and human (37) revealed conservation of intron number, but little similarity in their size or nucleotide sequence. All three taxa share six introns; however, their size differs dramatically. For example, intron 1 in tomcod and trout is 466 and 521 nt, respectively, whereas it is much larger in humans (2329 nt). Intron 4 is large in tomcod (414 nt), whereas it is much smaller in trout (194 nt) and humans (91 nt). When comparing intronic sequences among these taxa, similarity exceeding 70% was restricted to short 10-nt to 25-nt sequences.

In total, seven exons were identified. A consensus TATA box (TATAAA) was identified 25 nt upstream from the 5′ end of the CYP1A mRNA. The first exon does not contain any coding sequence; exons 2 through 7 contain the entire coding sequence. The putative start codon was located at nt 616–619 located in exon 2. The translational stop codon is 290 nt downstream of the 5′ end of the seventh exon. The remainder of the seventh exon contains 3′ noncoding sequences (842 nt). Consensus GT-AG RNA splice sites were observed at all exon–intron junctions.

Primer extension analysis was used to identify the transcriptional start site. Primer P12 was generated near the 5′ end of the cDNA sequence and employed in primer extension experiments. The major 92-bp extension product aligned with the adenine residue (Fig.


*
←P12
P1→
←P10
←P2
P3→
P6→
P8→
←P11
P9→
←P4

5) which is 13 bp upstream from the guanine residue at the 5′ end of the tomcod CYP1A cDNA (Fig. 2).

In this and earlier studies, Northern blot analyses of tomcod revealed variability in the number of CYP1A-hybridizable bands among individual fish from the Hudson River (Fig. 2A). All tomcod displayed a 2.5-kb mRNA, while variant individuals also exhibited an additional 1.9-kb hybridizable mRNA species. The genetic basis of this variation was confirmed in an earlier study by Southern blot analysis and in this study by PCR analysis using primers P3 and P4 of DNA from the individuals who displayed the variant mRNA band (Fig. 2B). All individuals who exhibited the truncated message also displayed a second smaller CYP1A DNA fragment.

To begin to evaluate the biological significance of the variant CYP1A transcript and to localize the site of this deletion, we compared the DNA sequences of the common and truncated CYP1A alleles. To initially locate the area of the deletion, RT-PCR analysis of poly(A)$^+$ RNA was used. Use of a combination of primers P3 and P4 indicated that the deletion was in the 3′ coding region between nt 1273 and nt 2414. The amplicon from an individual with the common genotype was 1140 bp, whereas individuals with the variant genotype also exhibited a second 530-bp fragment. Both PCR and Southern blot analyses indicated that the common and variant alleles were in equimolar concentrations in fish with the variant genotype. To determine whether the variant transcript was caused by a splicing defect, genomic DNA was amplified from individuals with both the common and variant genotypes. Fish with the common genotype displayed a 1325-bp fragment, whereas individuals with the variant genotype displayed 1325-bp and 720-bp DNA fragments (Fig. 2B). This difference in the size of the amplicons generated from cDNA and genomic DNA was consistent and indicated the presence of the 187-bp sixth intron in the amplicons from genomic DNA.

Variant fragments from RT-PCR and genomic DNA amplification were subsequently sequenced directly with primers P4 and P6. A comparison of the variant and common alleles indicated the absence of a 606-bp sequence extending from nt 1706 to nt 2311 in the truncated allele (Fig. 2C). This deletion was located entirely in the seventh exon, 3′ to the translational stop codon (5′ nt 1655). All other nucleotides sequenced in the variant allele were identical to those observed in the common allele.

3′ UTRs of both common and variant tomcod CYP1A mRNA were examined for formation of optimum secondary structures based on the minimum folding energy. In the common transcript, there are three major stem–loop structures between nt 2048 and nt 2485; their sizes were 78, 97, and 170 nt. In contrast, only one stem–loop structure of 194 nt was observed in the variant transcript. The poly(A) signal was 6 nt downstream of the single stem–loop structure in the variant transcript and it was within the most 3′ stem–loop structure in the common transcript.

Population studies were conducted to determine the frequency of the variant CYP1A allele in tomcod from six Atlantic coast rivers in the U.S. and Atlantic Canada. Both Southern blot analysis and PCR analysis using primers P3 and P4 were used to identify CYP1A genotypes. About 10% (23/268) of tomcod from the Hudson River exhibited this polymorphism; it was totally absent in tomcod from all other rivers. This difference in the frequencies of CYP1A1 genotypes between the Hudson River and other populations was significant ($\chi^2 = 10.6$, $P = 0.016$). Furthermore, the variant allele was only observed in a heterozygous state.

DISCUSSION

The Atlantic tomcod common CYP1A allele bears significant similarity to the trout, plaice, and mammalian CYP1A1 genes in terms of organization, nucleotide sequence, and amino acid sequence. For example, all CYP1A genes have seven exons and six introns and the relative length of each of the exons is very similar in all species. In all taxa, the length of the seventh (1050–1265 nt) and second (847–871 nt) exons far exceeds that of other exons. Lengths of exons 3 through 6 are almost identical in all species (127–128, 90, 124, and 87 nt, respectively). For comparison, the rabbit (38) and human (39) CYP1A1 allele also have seven exons and in both cases exons 2 and 7 are the largest. Amino acid sequence similarity is moderately high between tomcod and rainbow trout (71.7%) and between tomcod and human (52.2%). Although the number of CYP1A1 introns is conserved among tomcod, rainbow trout, and humans, their relative sizes and sequences showed little similarity (>70%) except for short sequences of less

FIG. 3. The nucleotide sequence of Atlantic tomcod CYP1A genomic DNA from a fish with the common genotype. Exons are indicated by bold lettering. Seven exons and six introns are indicated. Nucleotide positions of the cloned cDNA are indicated in parentheses. The deletion extending from nt 1709 to nt 2314 (cDNA), found in the 3′ UTR of the variant allele, is indicated by downward arrows. Sequence begins with the TATA box. The * at nt 31 (A) indicates the 5′ end of the CYP1A1 mRNA. Underlines at nt 616–619 indicate the putative start codon, nt 3494–3496 the stop codon, and nt 4316–4323 the poly(A) signal. Downward arrows indicate the start (nt 3547) and end of (nt 4152) the deletion found in individuals with the variant genotype. The locations of the nine primers used in PCR amplification or sequencing of the tomcod cDNA and genomic DNA are indicated. Direction of arrows at the 3′ end indicates the direction of synthesis by DNA polymerase. The GenBank accession numbers are L41886 and L41917.

```
          1                                                          60
Tomcod    MALMILPLIG SVSVSETLVA MITVCMIYML MKFLHPDVPE A.PPAPGPQA LPIIGNVWSW
Plaice    -M--M--F-- ---------- -T-M-LV-LI L--FQTEI.. GLRRL...KP .......LGL
Trout     -V-----I-- ---A--G--- -V-L-LV--I --YM-TEI-- GLKRL---KP -------LEV
Human     ...-LF-I.. -M-AT-F-L- SVIF-LVFWV IRASR-Q--K GLKNP---WG W-L--HMLTL

          61                                                         120
Tomcod    E.NGLPEPHG HGQRYGDILQ IQIGMRSVVV LSGHETVRQA LIKQGHDL.R PPDLYSFQFI
Plaice    GSKPYLSLTA MSK---HVF- ------P--- ---TG----- -----DEFAG -------R--
Trout     HN-PHLSLTA MSE---SVF- ------P--- ---SE----- -----E-FAG -------K--
Human     GK-PHLALSR MS-Q---V-- -R--STP--- ---LD-I--- -VR--D-FKG R----T-T--

          121                                                        180
Tomcod    NDGKSLAFST DQAGVWSPPK .....AGHEC PALLFHARGH HAAVLLHVGG ACLQGGRLPR
Plaice    -A-------- ------RARR KLAYS-LRSF ST-EGTTPEY SCVLEE-I.. .-KE-EY-IK
Trout     ---------- -K----RARR KLAMS-LRSF AT-EGTTPEY SC-LEE--.. .-KE-EY-VK
Human     SN-Q-MS--P -SGP--AARR RLAQNGLKSF SIASDP-SST SCYLEE--.. .SKEAEV-IS

          181                                                        240
Tomcod    QAAVQRHGTD ASFDPFRHIV VSVANVICGM CFGRRYGHED QELLSLVNLT DEFGKVVGSG
Plaice    -LNTVMKA.- G--------- ---------- ------D-D- ---V---T-S ----R-----
Trout     -LTSVMDV.S G--------- ---------- ------S-D- ----G---MS ----Q-----
Human     TLQELM.AGP GH-N-Y-YV- ---T----AI ------D-NH ---------N NN--E-----

          241                                                        300
Tomcod    NLADFIPLLR FLPNATMKRF MAINERFMTF VQKIVTEHYN TYDKDNIRDI TDSLIDHCED
Plaice    -P-----I-Q Y--S E--N- LR---H-TE- ---------T -FN------- ----------
Trout     -P--S--I-- Y---R----- -D--D--N-- --N-IS---E S-D------- ----------
Human     -P-----I-- Y---PSLNA- KDL--K-YS- M--M-K---K -FE-GH---- -----E--QE

          301                                                        360
Tomcod    RKLDENSNIQ MSDEKIVGIV NDLFGAGFDT VSTALSWSVM YLVAHPEIQE RLHQEIKDKV
Plaice    --------V- ---------- ---------- ---------- ---------- --Y---E---
Trout     ------A--- V--------- ---------- I------A-V ----Y----- -----L-E--
Human     KQ----A-V- L-----IN-- L--------- -T--I---L- ---MN-RV-R KIQE-LDTVI

          361                                                        420
Tomcod    GLSRSPVLTD RHNLPILEAF IFEIFRHSSF LPFTIPHCAT KDTSLNGYFI PKDTCVFINQ
Plaice    --D-M-L-S- KP---F---- -L--L----- --------T- ---------- ----------
Trout     -MI-T-R-S- KI---L---- -L-------- --------TI ---------- ----------
Human     -R--R-R-S- -SH--YM--- -L-T------ V------ST- R----K-FY- --GR---V--

          421                                                        480
Tomcod    WQINHDPELW KEPSTFNPDR FLSADASELN KLAGEKVMLF GMGKRRCIGE MVARNEVFLF
Plaice    ---------- -D--S----- -----G--V- --D-----A- ---------- VI-----Y--
Trout     --V----D-- ----S----- -----GT--- --E----LV- --D------- AIG-------
Human     ------QK-- VN--E-L-E- --TP-GA.ID -VLS---II- ---K-K---- TI--W-----

          481                                                526
Tomcod    LAILVQRLTF HAVPGEPLDM TPEYGLTMKH KRC.HLRATV RTTE
Plaice    ---II-K-H- LPI---K--- ---------- ---.--K--M -ARNEH
Trout     ----L---R- QEK--H---- ---------- ---.Q-K-SM -PWGQ
Human     ----L--VE- SVPL-VKV-- --I------- AC-E-FQMQL -S
```

FIG. 4. A comparison of the deduced amino acid sequence for Atlantic tomcod, plaice, rainbow trout, and human CYP1A1 genes. Identical amino acids are indicated (–). Gaps (.) were introduced to facilitate sequence alignment.

than 25 nt. In additional studies (Roy *et al.*, in preparation), we have demonstrated that the common CYP1A allele in tomcod shares identical 5′ dioxin response elements with those reported in mammalian CYP1A1 genes. In contrast, these enhancer elements are not observed upstream of mammalian CYP1A2 genes. Based on these results and the guidelines issued by the committee on nomenclature (40), we believe that this tomcod CYP1A allele should be assigned to the CYP1A1 subfamily.

In comparison to the common tomcod CYP1A1 allele, the variant allele exhibits a 606-bp deletion which is

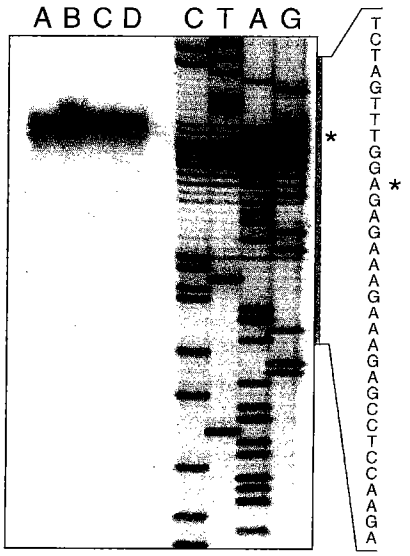


FIG. 5. Primer extension analysis to identify the 5′ end of tomcod CYP1A1 mRNA using total hepatic RNA and primer P12. Lanes A–D show reverse-transcriptase reaction products obtained by using RNA from four different fish. * indicates the 5′ end of the CYP1A1 mRNA. Sequencing reactions of tomcod genomic DNA using labeled primer P12 are shown in the four right lanes along with the actual nucleotide sequence.

located entirely within the seventh exon (cDNA nt 1706–2311). The deletion begins 51 nt beyond the translational stop codon and thus is located within the 3′ untranslated region of the tomcod CYP1A1 mRNA. With the exclusion of this deletion, the common and variant CYP1A1 alleles share identical nt sequences. Identity of the two sequences confirms that this variant is a CYP1A1 allele and furthermore suggests that this polymorphism has arisen in the very recent evolutionary past. Population studies revealed that the variant CYP1A1 allele is restricted to the Hudson River population. It was absent in tomcod from the other populations that were screened in this study. Of course, tomcod populations exist in many other estuaries between New York and southern Maine, and these were not screened for CYP1A1 genotypes. It is possible that the elevated frequency of the CYP1A1 variant allele in the Hudson River population is a result of stochastic population processes unrelated to xenobiotic exposure. This possibility could be initially addressed by screening of tomcod populations between Maine and the Hudson River for this polymorphism. However, the fact that no individuals exhibited a homozygous variant genotype suggests that selection against the variant allele may be operative in the Hudson River population. We are currently evaluating this possibility by crossing tomcod with heterozygous CYP1A1 genotypes and challenging their offspring with model xenobiotics.

Several lines of evidence suggest that the variant allele is not a pseudogene: (i) it is transcribed; (ii) sequences of the variant and common alleles are identical. It is unlikely that sequence conservation between a pseudogene and its ancestral gene would be rigidly maintained. (iii) DNA concentrations in the common and variant alleles in a heterozygous fish are equimolar.

Several genetic polymorphisms in CYP1A1 have been described in human populations and they have been implicated as markers of genetic susceptibility to lung cancer in Asians (41) and African Americans (42), but not in Caucasian populations (43). In some instances, these CYP1A1 genetic polymorphisms have been correlated with effects on levels of CYP1A1 mRNA (44), protein (45), and enzymatic activity (46). In all instances, variant alleles contained single base substitutions either in the coding region of exon 7 (47) or in 3′ noncoding regions (48). It is likely that substitutions proximal to the CYP1A1 gene's catalytic site in exon 7 are of more functional significance than in those 3′ of exon 7. Large deletions comparable to those we report in tomcod have yet to be observed in mammalian species. Thus, it would appear that the impact of these polymorphisms on susceptibility to neoplastic disease is dependent on (i) the location of the polymorphism, (ii) the genetic background of the populations which harbor the polymorphism, and (iii) the tissues affected.

In comparison to these studies in humans, the CYP1A1 genetic polymorphism in tomcod is in exon 7, yet 3′ to the translational stop codon. Thus, it is unlikely that this polymorphism impacts on structure or enzymatic activity of the CYP1A1 protein. Studies which directly compare CYP1A1 enzymatic activity in tomcod with the normal and variant alleles are needed. However, the fact that this polymorphism is restricted to the cancer-prone Hudson River population is intriguing and suggests a possible link between this polymorphism and susceptibility to neoplasia. Equally intriguing is the size of the deletion, its position, and the growing body of literature on the functional significance of the 3′ untranslated regions of genes (49) in processes such as differentiation (50) and tumor suppression (51). A recurrent theme from these studies is the importance of 3′ UTR sequences on the stability of mRNA species and therefore regulation of gene expression.

It has been demonstrated that mRNA stability differs greatly among transcripts. For immediate early genes, such as c-jun, c-myc, and c-fos, transcript half-lives may be less than 30 min (52), whereas in structural genes such as vitellogenin half lives persist for 500 h (53). Recent studies have demonstrated that binding of regulatory proteins to 3′ UTR mRNA sequences serves to stabilize transcripts (54) or alternatively to hasten their degradation by triggering deadenylation (55) or serve as an anchor for endonuclease

cleavage (56). Specific 3′ UTR motifs such as AUUUA pentanucleotides (55), U-rich sequences (52), and stem–loop structures (57), have been identified with which these inducible proteins interact.

In tomcod, we have detected differences in the concentrations of common and variant CYP1A1 mRNA in fish treated with model inducers such as B[*a*]P, suggesting that the stability of the two transcripts may differ. Furthermore, computer simulations suggest the presence of three stem–loop structures in the 3′ UTR of the common tomcod CYP1A1 mRNA, but only one stem–loop structure in the variant transcript. Additionally, two AUUUA pentanucleotide motifs were identified in the tomcod 3′ UTR. The presence of these potential protein binding nucleotide motifs and secondary structures in the CYP1A1 3′ UTR suggests the possible significance of these landmarks in processing of CYP1A1 mRNA. Furthermore, the locations of these potentially interactive sequences within or in close proximity to the deletion suggest a mechanism whereby stability of common and variant CYP1A1 mRNA may differ.

ACKNOWLEDGMENTS

We thank Lorraine Maceda, Jannine Rosko, and Kate Van Cleef for technical assistance, Simon Courtenay, Mark Mattson, Dennis Duning, and Julian Dodson for sample acquisition, and the Hudson River Foundation, NIH Grants ES05003 and ES05541, and NIEHS Center Grant ES00260 for support of this study.

REFERENCES

1. McMahon, G. (1994) *Environ. Health Perspect.* **102**(Suppl. 12), 75–80.
2. Malins, D. C., McCain, B. B., Landahl, J. T., Myers, M. S., Krahn, M. M., Brown, D. W., Chan, S.-L., and Roubal, W. T. (1988) *Aquat. Toxicol.* **11,** 43–67.
3. Baumann, P. C., and Harshbarger, J. C. (1995) *Environ. Health Perspect.* **103,** 168–170.
4. Johnson, L. L., Stehr, C. M., Olson, O. P., Myers, M. S., Pierce, S. M., Wigren, C. A., McCain, B. B., and Varanasi, U. (1993) *Environ. Sci. Technol.* **27,** 2759–2771.
5. Myers, M. S., Stehr, C. M., Olson, O. P., Johnson, L. L., McCain, B. B., Chan, S.-L., and Varanasi, U. (1994) *Environ. Health Perspect.* **102,** 200–215.
6. Dey, W. P., Peck, T. H., Smith, C. E., and Kreamer, G.-L. (1993) *Can. J. Fish. Aquat. Sci.* **50,** 1897–1907.
7. Cormier, S. M., and Racine, R. N. (1990) *in* Biological Markers of Environmental Contamination (McCarthy, J. F., and Shugart, L. R., Eds.), pp. 59–71, Lewis, Chelsea, MI.
8. Squibb, K. S., O'Connor, J. M., and Kneip, T. J. (1991) New York/New Jersey Harbor Estuary Program Module 3.1: Toxics Characterization Report, Institute of Environmental Medicine, New York University Medical Center, Tuxedo, New York.
9. National Status and Trends Program for Marine Environmental Quality (1988) Progress Report: A Summary of Selected Data on Chemical Contaminants in Sediments Collected during 1984, 1985, 1986, and 1987, NOAA Technical Memorandum NOS OMA 44, Rockville, MD, National Oceanic and Atmospheric Administration.
10. Rappe, C., Bergqvist, P.-A., Kjeller, L.-O., and Swanson, S. (1991) *Chemosphere* **22,** 239–266.
11. Klauda, R. J., Peck, T. H., and Rich, G. K. (1981) *Bull. Environ. Contam. Toxicol.* **27,** 829–835.
12. Kreamer, G.-L., Squibb, K., Gioeli, D., Garte, S. J., and Wirgin, I. (1991) *Environ. Res.* **55,** 64–78.
13. Wirgin, I. I., Grunwald, C., Courtenay, S., Kreamer, G.-L., Reichert, W. L., and Stein, J. E. (1994) *Environ. Health Perspect.* **102,** 764–770.
14. Wirgin, I., Currie, D., and Garte, S. J. (1989) *Carcinogenesis* **10,** 2311–2315.
15. Wirgin, I., Konke, B., Pedersen, M., and Grunwald, C. (in press) *Estuaries.*
16. Courtenay, S., Williams, P. J., Grunwald, C., Konkle, B., Ong, T.-L., and Wirgin, I. I. (1994) *Environ. Health Perspect.* **102**(Suppl. 12), 85–90.
17. Stegeman, J. J., and Hahn, M. E. (1994) *in* Aquatic Toxicology: Molecular, Biochemical, and Cellular Perspectives (Malins, D. C., and Ostrander, G. K., Eds.), pp. 87–206, Lewis, Boca Raton, FL.
18. Varanasi, U., Reichert, W. L., and Stein, J. E. (1989) *Cancer Res.* **49,** 1171–1177.
19. Maronpot, R. R., Foley, J. F., Takahashi, K., Goldsworthy, T., Clark, G., Tritscher, A., Portier, C., and Lucier, G. (1993) *Environ. Health Perspect.* **101,** 634–642.
20. Courtenay, S., Grunwald, C., Kreamer, G.-L., Alexander, R., and Wirgin, I. (1993) *Aquat. Toxicol.* **27,** 225–244.
21. Wirgin, I., Kreamer, G.-L., and Garte, S. J. (1991) *Aquat. Toxicol.* **19,** 205–214.
22. Wirgin, I., D'Amore, M., Grunwald, C., Goldman, A., and Garte, S. J. (1990) *Biochem. Genet.* **8,** 459–475.
23. Southern, E. M. (1975) *J. Mol. Biol.* **98,** 503–517.
24. Wahl, G. M., Stern, M., and Stark, G. R. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 3683–3687.
25. Rigby, P. W. J., Dieckmann, M., Rhodes, C., and Berg, P. (1977) *J. Mol. Biol.* **113,** 237–251.
26. Heilmann, L. J., Sheen, Y.-Y., Bigelow, S. W., and Nebert, D. W. (1988) *DNA* **7,** 379–398.
27. Chomczynski, P., and Sacchi, N. (1987) *Anal. Biochem.* **162,** 156–159.
28. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
29. Feinberg, A. P., and Vogelstein, B. (1983) *Anal. Biochem.* **132,** 6–13.
30. Turner, D. H., Sugimoto, N., and Freier, S. M. (1988) *Annu. Rev. Biophys. Biophys. Chem.* **17,** 167–192.
31. Roff, D. A., and Bentzen, P. (1989) *Mol. Biol. Evol.* **6,** 539–545.
32. Wirgin, I. I., Kreamer, G.-L., Grunwald, C., Squibb, K., Garte, S. J., and Courtenay, S. (1992) *Mar. Environ. Res.* **34,** 103–108.
33. Leaver, M. J., Pirrit, L., and George, S. G. (1993) *Mol. Mar. Biol. Biotechnol.* **2,** 338–345.
34. Berndston, A. K., and Chen, T. T. (1994) *Arch. Biochem. Biophys.* **310,** 187–195.
35. Jaiswal, A. K., Gonzalez, F. J., and Nebert, D. W. (1985) *Science* **228,** 80–83.
36. Gonzalez, F. J., Kimura, S., and Nebert, D. W. (1985) *J. Biol. Chem.* **260,** 5040–5049.
37. Kawajiri, K., Wantanabe, J., Gotoh, O., Tagahira, Y., Sogawa, K., and Fujii-Kuriyama, Y. (1986) *Eur. J. Biochem.* **159,** 219–225.

38. Strom, D. K., Postlind, H., and Tukey, R. H. (1992) *Arch. Biochem. Biophys.* **294,** 707–716.
39. Iverson, P. L., Heiger, W. J., Bresnick, E., and Hines, R. N. (1987) *Arch. Biochem. Biophys.* **256,** 397–401.
40. Nelson, D. R., Kamataki, T., Waxman, D. J., Guengerich, F. P., Estabrook, R. W., Feyereisen, R., Gonzalez, F. J., Coon, N. J., Gunsalus, I. C., Gotoh, O., Okuda, K., and Nebert, D. W. (1993) *DNA Cell Biol.* **12,** 1–51.
41. Hayashi, S.-I., Watanabe, J., and Kawajiri, K. (1992) *Jpn. J. Cancer Res.* **83,** 866–870.
42. Taoli, E., Crofts, F., Trachman, J., Demopoulos, R., Toniolo, P., and Garte, S. J. (1995) *Cancer Res.* **55,** 472–473.
43. Hirvonen, A., Husgafvel-Pursiainen, K., Karjalainen, A., Anttila, S., and Vainio, H. (1992) *Cancer Epidemiol. Biomarkers Prev.* **1,** 485–489.
44. Cosma, G., Crofts, F., Currie, D., Wirgin, I., Toniolo, P., and Garte, S. J. (1993) *Cancer Epidemiol. Biomarkers Prev.* **2,** 53–57.
45. Crofts, F., Taioli, E., Trachman, J., Cosma, G. N., Currie, D., Toniolo, P., and Garte, S. J. (1994) *Carcinogenesis* **15,** 2961–2963.
46. Petersen, D. D., McKinney, C. E., Ikeya, K., Smith, H. H., Bale, A. E., McBride, O. W., and Nebert, D. W. (1991) *Am. J. Hum. Genet.* **48,** 720–725.
47. Hayashi, S.-I., Watanabe, J., Nakachi, K., and Kawajiri, K. (1991) *J. Biochem.* **110,** 407–411.
48. Crofts, F., Cosma, G. N., Currie, D., Taioli, E., Toniolo, P., and Garte, S. J. (1993) *Carcinogenesis* **14,** 1729–1731.
49. Atwater, J. A., Wisdom, R., and Verma, I. (1990) *Annu. Rev. Genet.* **24,** 519–541.
50. Rastinejad, F., and Blau, H. M. (1993) *Cell* **72,** 903–917.
51. Rastinejad, F., Conboy, M. J., Rando, T. A., and Blau, H. M. (1993) *Cell* **75,** 1107–1117.
52. You, Y., Chen, C.-Y. A., and Shyu, A.-B. (1992) *Mol. Cell. Biol.* **12,** 2931–2940.
53. Dodson, R. H., and Shapiro, D. J. (1994) *Mol. Cell. Biol.* **14,** 3130–3138.
54. Zaidi, S. H. E., and Malter, J. S. (1994) *J. Biol. Chem.* **269,** 24007–24013.
55. Shyu, A.-B., Belasco, J. G., and Greenberg, M. E. (1991) *Genes Dev.* **5,** 221–231.
56. Binder, R., Horowits, J. A., Bailion, J. P., Koeller, D. M., Klausner, R. D., and Harford, J. B. (1994) *EMBO J.* **13,** 1969–1980.
57. Williams, A. S., and Marzuff, W. F. (1995) *Nucleic Acids Res.* **23,** 654–662.