Isolation and characterization of two distinct growth hormone cDNAs from the tetraploid smallmouth buffalofish (Ictiobus bubalus)

Mark D. Clements, a,b Henry L. Bart Jr., a,b and David L. Hurley c,*

a Department of Ecology and Evolutionary Biology, Tulane University, New Orleans, LA, USA
b Tulane University Museum of Natural History, Belle Chasse, LA, USA
c Department of Biochemistry, Tulane University School of Medicine, 1430 Tulane Avenue, New Orleans, LA 70112-2699, USA

Received 2 December 2003; revised 8 January 2004; accepted 13 January 2004

Abstract

The growth hormone (GH) gene has been characterized for a number of fishes and used to establish phylogenetic relationships and population structures. Analysis of tetraploid fishes, such as salmon and some Asian cyprinids, has shown the presence of two GH genes. Fishes in the sucker family (Catostomidae, Cypriniformes) are also tetraploid, and the present study reports the isolation and characterization of two GH cDNAs from a representative species, the smallmouth buffalofish (Ictiobus bubalus). The GH cDNAs of smallmouth buffalofish are 1272 and 1273 nt in length, and each codes for a polypeptide of 210 amino acids, predicted to be cleaved to a final product of 188 aa. The GH cDNAs of smallmouth buffalofish are 6% divergent in nt sequence in the coding region, and there are 16 differences in predicted aa sequence. Because the cDNAs have distinct sequences in coding regions and in UTRs, which differed by more than 10%, they were identified as GHI and GHII. The predicted GHI protein contains 4 Cys residues, homologous to other vertebrate GH sequences. On the other hand, GHII has 5 Cys residues, homologous to other ostariophysan sequences. GHI and GHII are most similar to other cypriniform fishes for both nt and protein sequences. Phylogenetically, the sequences of smallmouth buffalofish GH consistently grouped with Asian cyprinids, but not loaches, consistent with morphological evidence suggesting that suckers are most closely related to minnows.

© 2004 Elsevier Inc. All rights reserved.

Keywords: Phylogenetics; Comparative genomics; Bootstrap analysis; DNA sequencing; Endocrinology

1. Introduction

Growth hormone (GH) is a single-chain, pituitary-specific hormone that is essential for promotion, and maintenance of somatic growth in vertebrates (Chen et al., 1994; Yowe and Epping, 1995, 1999). Because GH is present and functionally constrained within a large number of taxa, comparison of GH sequences has been informative in a variety of phylogenetic analyses. For example, GH sequences amongst teleost fishes have been used to clarify evolutionary relationships (Rubin and Dores, 1994; Rubin et al., 1996; Schneider et al., 1992; Venkatesh and Brenner, 1997). Beyond the use of the sequences for phylogenetic analysis, there are practical applications, such as increasing efficiency and yield in aquaculture (Almuly et al., 2000).

GH sequences have been used to study evolutionary relationships of fishes at a variety of taxonomic levels. GH coding sequences have been used to resolve phylogenetic relationships of major clades of fishes, and the results have been validated because of high agreement with relationships based on morphology and other data (Almuly et al., 2000; Bernardi et al., 1993; Rubin and Dores, 1994, 1995; Rubin et al., 1996). GH intron sequences have also been used to infer sub-familial phylogenetic relationships of salmonids (Oakley and Phillips, 1999) and to characterize intraspecific, population genetic structures of some fishes (Almuly et al., 2000; Schlee et al., 1996; Yowe and Epping, 1995; Yue et al., 2001). In this way, the GH gene is extremely informative in resolving both recent and distant divergence.
To develop amplification primers for use in family-level phylogenetic analyses of catostomids, GH cDNAs from *Ictiobus bubalus* were characterized. Because catostomids are tetraploid (Ferris and Whitt, 1977; Uyeno and Smith, 1972), *I. bubalus* was expected to have two paralogous cDNAs as seen in other tetraploids, such as salmonids (Oakley and Phillips, 1999) and cyprinids (Law et al., 1996). Therefore, it was necessary to develop locus-specific primers by first amplifying and sequencing GH cDNAs from *Ictiobus bubalus* pituitary RNA. These sequences must be obtained to establish sequence polymorphisms between the two paralogous copies of GH that allow for locus-specific amplification of genomic DNA. Because GH sequences, especially coding sequences, appear well suited for resolving phylogenetic relationships among major clades of fishes, the addition of GH sequences for catostomids will likely clarify basal relationships of Cypriniformes, particularly the sister group of catostomids (Harris and Mayden, 2001).

2. Results and discussion

2.1. Isolation of two GH cDNAs from *I. bubalus*

Using a pair of degenerate primers designed to anneal to conserved sequences in an alignment of all cypriniform fish GH cDNAs present in GenBank and spanning the central portion of the GH coding region (Fig. 1A), amplification products were obtained from *I. bubalus* pituitary cDNA (Fig. 1B) in the predicted size range. These products were then cloned and sequenced to determine this portion of the coding region for putative GH cDNA of *I. bubalus*. The isolates analyzed consistently yielded two different nucleotide sequences, and BLAST searching of the NCBI database indicated that both were similar to other teleost GH cDNAs. Further characterization of these cDNAs employed an RNA ligase mediated RACE method to determine 5' and 3' ends using amplification with a specific UAP and GH primer (Fig. 1A). After obtaining 5' and 3' UTR sequences, which again yielded two consistent, but different sequences, the full coding region sequence of the two cDNAs were selectively amplified using locus-specific primers (Fig. 1B).

2.2. Sequences of GHI and GHII cDNAs

The complete sequences of the full length cDNAs were determined by using the different amplification products covering the entire GH cDNA from *I. bubalus* (Fig. 2). Because the presence of two copies of GH in this species is consistent with the tetraploid genome of catostomids (Uyeno and Smith, 1972), and because the cDNAs have distinct sequences, the two cDNA forms obtained are indicative of two paralogous copies of GH. These sequences will be useful for the development of species-specific primers for use in phylogenetic analyses of catostomids.

Fig. 1. Fish GH cDNA amplification primer design and products. (A) The length of the GH mRNA of 5' UTR, coding region (boxes), and 3' UTR. Primers designed for amplification of *I. bubalus* GH are shown as arrows originating at the 5' nt of each primer and pointing in the 3' direction. (B) The amplification products obtained from whole pituitary cDNAs in PCR reactions with primers: GHI and GHR, lane 1; GSPF and 3' UAP, lane 2; GSPR and 5' UAP, lane 3; GHIIF and 3' UTR, lane 4; and GHIIF and 3' UTR; lane 5. Lane M contains 1 kb Plus (Invitrogen) markers of known molecular weights given in bp next to the fragments.
sequences were identified as GHI (GenBank Accession No. AY375301) and GHII (GenBank Accession No. AY375302). The total length of each cDNA is 1272 nt for GHI and 1273 nt for GHII. Both cDNAs have 5' UTRs of 56 nt and open reading frames of 633 nt. GHI has 583 nt of 3' UTR, while GHII has a 584 nt 3' UTR.

Fig. 2. *I. bubalus* GHI (GenBank Accession No. AY375301) and GHII (GenBank Accession No. AY375302) cDNA nucleotide and amino acid sequences. The complete nucleotide sequence of the GHI cDNA is listed as the middle entry in the figure. Below the GHI sequence, nucleotide differences in GHII cDNA are given at the position where they occur. A dash represents nucleotide identity between the two GH cDNAs. Above the predicted codons of the GHI nucleotide sequence is the amino acid sequence for GHI protein. Below the GHII nucleotide sequence is given any amino acid in the GHII protein that differ from GHI. Regions without amino acids associated with them are either the 5' or 3' ends of the mRNA. In these regions, * indicate the presence of insertions added to maximize alignment homology. The polyadenylation site is underlined in both GHI and GHII, and the polyA tail is indicated with (An).
Each cDNA has a standard polyadenylation signal, ATTTAA, in the 3' region located at a typical distance upstream from the poly-A site (Chiou et al., 1990; Koren et al., 1989; Law et al., 1996). Pairwise I. bubalus GHI and GHII sequence comparison revealed 8 nt differences in the 5' UTR and 58 nt differences in the 3' UTR. As a percentage, the 14% divergence of the 5' UTRs is greater than the 10% difference of the 3' UTR. However, 8 gaps are needed to obtain maximum homology at the 3' end (Fig. 2), and may invalidate the percentage comparison.

The predicted aa sequences encoded by GHI and GHII were then analyzed from these cDNAs. Both GHI and GHII encoded for a protein of 210 aa, the same length as other cypriniform GH proteins (Ho et al., 1991; Hong and Schartl, 1993; Koren et al., 1989). Both cDNAs also have a signal peptide cleavage site at Ser23 that is found in other cypriniform sequences and thus mature peptides of 188 aa are predicted for both cDNAs (Noh et al., 1999). In the coding region, the two cDNAs are 6% divergent in nt sequence: there are 22 silent substitutions in codons and 18 that alter the protein sequence. However, because 4 of these are multiple changes within the same two codons, the final result is a total of 16 differences in the deduced aa sequences between GHI and GHII.

2.3. Comparison of GHI and GHII to other teleost sequences

In the coding region, I. bubalus GH cDNAs are most like those of other cypriniform fishes for both nt (~78% similarity) and aa (~86% similarity) comparisons. The I. bubalus cDNAs are more divergent from other, more derived teleost sequences; on average, the GH genes of smallmouth buffalofish are 28% divergent from protoacanthopterygians and 42% divergent from percomorphs in nt sequence, and the proteins are 32 and 47% divergent from these respective groups. Rubin et al. (1996) identified 44 invariant aa residues in a comparison of teleostean GH sequences available at that time. All of these aa residues are also conserved in smallmouth buffalofish GHI and GHII cDNAs. Because these invariant residues are located within the five conserved domains identified by Watahiki et al. (1989), and are thought to be necessary for proper functioning of GH, it is predicted that the both GH cDNAs of I. bubalus encode functional hormones.

These two I. bubalus GH coding sequences have diverged subsequent to duplication by 7.6% at the amino acid and 6.1% at the nucleotide level. In comparison, tetraploid salmonid GH gene divergences range from 2.8 to 7.1% for aa sequences and 3.7 to 5.3% for nt sequences, and tetraploid cyprinid GH gene divergences range from 3.3 to 5.2% for aa sequences 4.4 to 6.6% for nt sequences. Thus, because divergence of I. bubalus GHI and GHII falls within the range of divergence seen for GH sequences in these fishes, it is likely that all these paralogous GH gene copies have diverged at similar rates subsequent to duplication. The similar rate in GH gene divergence might be due to several factors. One is that the genome duplication events have similar ages, as tetraploidy events for these fishes are estimated to have occurred 25–100 million years ago (mya) for salmonids (Agellon et al., 1988), 16–50 mya for tetraploid cyprinids, and 50 mya for catostomids (Larhammar and Rissing, 1994). Another reason for similar divergence may be functional constraints acting on duplicate loci of GH in these fishes because both copies are required for normal physiology. Such functional constraints are suggested by the fact that both duplicate loci of GH are expressed in pituitary mRNA, a primary source for isolation of these cDNAs (Law et al., 1996; Yang et al., 1997; and the present report). Interestingly, it has been shown that GHI and GHII mRNA abundance in rainbow trout is not equal (Yang et al., 1997), but possible selective roles of GHI and GHII gene products are unknown.

Paired Cys residues are thought to be particularly important for proper folding and thus three dimensional structure of GH. As in other vertebrates, GHI of I. bubalus contains 4 conserved Cys residues (C71, C183, C200, and C208). Conversely, GHII has 5 Cys residues with an additional Cys at position 145. C145 is homologous to additional, unpaired Cys residues in other ostariophysan sequences (Anathy et al., 2001; Ho et al., 1991; Hong and Schartl, 1993; Koren et al., 1989). Law et al. (1996) also discovered two GH cDNA variants in goldfish, Carassius auratus (a tetraploid ostariophysan), that have either 4 or 5 Cys residues. Little is known about the functional significance of the extra Cys residue. One study showed that in common carp (another tetraploid ostariophysan), 5 Cys residue-containing GH proteins were more likely to form oligomeric complexes, perhaps because the extra Cys interfered with disulfide bonding and proper folding of GH (Fine et al., 1993).

2.4. Phylogenetic analysis of GHI and GHII

A phylogenetic tree was constructed in order to define the relationship of these two, presumably paralogous, GH genes in I. bubalus to other fishes. As shown in Fig. 3, this resulted in a consensus phylogeny that resolves major clades of fishes and is consistent with relationships proposed by morphology, and other data (Bernardi et al., 1993; Nelson, 1994; Rubin and Dores, 1995). Ostariophysans are monophyletic and sister to protoacanthopterygians (salmonids and allies) plus percomorphs. Within Cypriniformes, louches (cobitids) are basal and sister to cyprinids plus catostomids. Interestingly, our phylogenetic analysis suggests that suckers evolved from a cyprinid-like ancestor, but tet-
raploidy events in catostomids and Asian cyprinids then occurred independently, as neither GHI nor GHII sequences of smallmouth buffalofish, common carp, or goldfish are monophyletic. The phylogenetic grouping of catostomids with cyprinids, rather than cobitids, is consistent with the evolutionary hypotheses of Smith (1992) and Uyeno and Smith (1972), but differs from that of Harris and Mayden (2001).
In summary, it has been shown that the *I. bubalus* produces two GH cDNAs that differ considerably in nucleotide and predicted protein sequence, a result consistent with the tetraploidy of the catostomid fishes. The nucleotide and amino acid sequences of *I. bubalus* GHI and GHII are most similar to those of other Cypriniformes, with perfect identity at the 44 conserved amino acids invariant in other teleost fishes. The differences between GHI and GHII are similar to those of the two GH proteins in goldfish, particularly the different number of Cys residues in each specific gene. These GH cDNA sequences afford resolution of relationships among fishes in a manner consistent with morphology and suggest that suckers (Catostomidae) are sister to minnows, but not loaches.

### 3. Methods

#### 3.1. RNA extraction and reverse transcription

Adult *I. bubalus* were captured using a boat-electrofisher from the Amite River, Lake Pontchartrain drainage, LA. Fresh pituitaries were dissected from specimens according to Tulane University IACUC-approved protocols, frozen on dry ice and stored at −70°C until use. Total RNA was extracted using a modified acid guanidinium thiocyanate–phenol–chloroform protocol (Ultraspex, Biotech, Houston, TX). RNA was resuspended in TE and yield was quantified by UV spectrometry (Gene Quant, Pharmacia, Piscataway, NJ). First strand cDNA was reverse-transcribed from total RNA using oligo-dT12–18 and Superscript II RT (Invitrogen, Carlsbad, CA).

#### 3.2. PCR amplification and sequencing

A pair of degenerate primers (GHF: 5'-TCAGANAACCC-AGMGGCCTTCTC-3' and GHR: 5'-CAGG GTRCAGTTKGAATCSAR-3') were designed from alignments of all cypriniform GH sequences deposited in the GenBank as of November 2002. Primers were designed with the computer software Oligo 6.6 (MBI, Cascade, CO) and synthesized by Integrated DNA Technologies (Coralville, IA). PCRs were performed in 25 µl total volumes containing 1.5 mM MgCl2, 1x PCR buffer (Invitrogen), 0.2 mM each dNTP, 20 pmol of each primer, 0.5 U Platinum Taq DNA polymerase (Invitrogen), and 2 µl first strand cDNA. Amplification used a thermal profile consisted of an initial 94°C for 2 min, then 30 cycles of 94°C for 30 s, 52°C for 30 s, and 72°C for 1 min on an Eppendorf MasterCycler (Westbury, NY). PCR products were separated by electrophoresis in 1% agarose, stained with ethidium bromide, and visualized with a UV transilluminator–gel documentation system (Bio-Rad, Richmond, CA). Desired PCR products were isolated by gel elution (Qiagen, Valencia, CA) and ligated into pCR4-TOPO vectors. These were transformed into TOP 10 competent cells (TOPO TA Cloning Kit, Invitrogen) and plated on selective agar plates containing 50 µg/ml Kanamycin (Sigma, St. Louis, MO). Positive colonies were grown in overnight cultures and plasmid DNA extracted using affinity/elution methods (QIAprep kit, Qiagen). Insert size was determined by agarose gel electrophoresis after digestion with EcoRI in comparison to the 1 kb-Plus ladder (Invitrogen). Plasmid DNA from 15 to 20 isolates were sequenced using dye terminator cycle sequencing (Applied Biosystems) as described previously (VanderHeyden et al., 2000). After 30 cycles, excess dye terminators, primers, and nucleotides were removed by gel filtration (Edge Biosystems, Bethesda, MD) prior to analysis on an ABI 373A instrument using 6% polyacrylamide gels in 7 M urea (Sooner Scientific, Garvin, OK).

#### 3.3. Rapid amplification of cDNA ends

RNA ligase mediated RACE (Gene Racer Kit, Invitrogen) was used to isolate 5' and 3' ends of GH cDNAs according to the manufacturer’s protocol. Total RNA was treated with calf intestinal phosphatase, dephosphorylated with tobacco pyrophosphatase, and a GeneRacer RNA oligonucleotide sequence was ligated to the 5' end using T4 RNA ligase. Modified mRNA was then reverse transcribed with GeneRacer oligo-dT primer and Superscript II RT (Invitrogen) to provide cDNA with a known upstream universal adapter primer (5' UAP) in the GeneRacer oligonucleotide and a GH-specific downstream primers selected to anneal to both cDNAs from *I. bubalus* sequences determined. The 3' region of GH was amplified using primer GSFR (5'-GCTGACACCTCTGCATAG-3') and UAP (GR3: 5'-GCTGACACCTCTGCATAG-3'), and the 5' region was amplified using GSFR (5'-CRYTGATGCCCATGATATG-3') and a 3' UAP supplied by the manufacturer (GR5: 5'-CGACTGAGCCAGTGGCACTGA-3'). PCR was performed as described except that reaction volume was increased to 50 µl, MgCl2 concentration was increased 2.0 mM, and the annealing temperature was 59°C. After amplification for 35 cycles, RACE PCR products were cloned and then 15–20 clones were sequenced.

#### 3.4. Amplification of complete coding region sequences

Polymorphisms identified in 5' untranslated regions (UTRs) of *I. bubalus* GH cDNAs were used to design locus-specific primers to amplify complete coding sequences from oligo dT-primed first strand cDNA. The
locus specific primers GHHF (5'-AAAGCCTTACAATAGAC-TAAC-3') and GHHIF (5'-CAAACTTCAACTAGACTC-3') were used in conjunction with a 3'UTR primer annealing to either GHI or GHHI (GH3UTR: 5'-YACARGCTTGGCT-GCTCTTT-3') located approximately 50 bp beyond the putative stop codon. PCR was performed as described except that the annealing temperature was 58 °C and extension times were decreased to 30 s. PCR products were directly sequenced after purification with the QIAquick PCR purification Kit (Qiagen).

3.5. DNA sequence analysis

Raw sequence chromatograms were assembled into full length sequences and edited using Sequencher 4.1 (Gene Codes, Madison, WI) on a Macintosh G3 computer with OS 8.6 (Apple Computer). Nucleotide sequences were translated to amino acids and analyzed with MacVector 7.0 (Oxford Molecular). Pairwise and multiple alignments of nucleotide and amino acid sequences were performed using CLUSTAL W in the computer software MacVector 7.0. BLAST was accessed via the NCBI website at www.ncbi.nlm.gov. Substitution analysis, sequence divergence, and phylogenetic reconstructions were performed with PAUP b10 (Swofford, 2000).

Acknowledgments

We appreciate the editorial comments of Rebecca E. Blanton. This work was supported by NSF Grants IBN 9600805 to D.L.H. and NSF 0212991 to H.L.B.

References


