Zebrafish *ftz-f1* gene has two promoters, is alternatively spliced, and is expressed in digestive organs

Wen-wen LIN*, Hsin-Wei WANG*†, Cristina SUM*, Dong LIU‡, Choy L. HEW‡ and Bon-chu CHUNG*¹

*Institute of Molecular Biology, Academia Sinica, Nankang, Taipei, Taiwan, Republic of China, †Graduate School of Zoology, National Taiwan University, Taipei, Taiwan, Republic of China, and ‡Division of Structural Biology and Biochemistry, Hospital for Sick Children, Toronto, Ontario, Canada M5G 1X8

Fushi-tarazu Factor-1 (FTZ-F1) is a family of nuclear receptors involved in various developmental processes. We have cloned a zebrafish FTZ-F1 gene, termed *ff1*, which belongs to the fetoprotein transcription factor/liver receptor homologue-1 (FTF/LRH-1) subgroup of the FTZ-F1 family. Four transcripts arise as a result of differential promoter usage and alternative splicing at the 3'-most exons. The longer transcript, form A, encodes a transcriptional activator. The shorter transcript, form B, lacks the activation domain, and hence could not activate transcription. The difference in promoter usage generates FF1 proteins with different N-terminal sequences. All four transcripts appear to be

expressed in most of the adult tissues, whereas, during embryo development, the IIA form is the predominant transcript. Reverse transcriptase-PCR and *in situ* hybridization experiments showed that the *ff1* transcript is expressed in the hypothalamus, spinal cord, mandibular arch and digestive organs, including pancreas, liver, and intestine. The expression of *ff1* in the digestive organs implies its function in gut development.

Key words: fetoprotein transcription factor (FTF), gut, liver receptor homologue (LRH), nuclear receptor, steroidogenic factor 1 (SF1).

INTRODUCTION

Fushi-tarazu Factor-1 (FTZ-F1) was first found to be the transcription factor that activates the *ftz* gene in *Drosophila*, and was subsequently named as such [1,2]. Many FTZ-F1 homologues have been found in mouse, human, zebrafish and *Xenopus* [3–7]. These FTZ-F1 homologues contain a conserved sequence adjacent to the C-terminal end of the zinc-finger motif, termed the FTZ-F1 box [2]. In addition, they belong to the nuclear receptor superfamily, most members of which can activate transcription when bound to their ligands [8]. FTZ-F1 belongs to the orphan receptor family, which does not have known ligands, but can activate transcription [9,10]. Unlike other members in the nuclear receptor superfamily, which often dimerize and recognize the GGTC sequence as direct or inverted repeats, FTZ-F1 binds to the monomeric (A/T)CAAGG(A/T)C(C/G) sequence [11]. Many nuclear receptors activate transcription via an activation domain, AF-2, located at their C-terminus. Some of the FTZ-F1 members do not have this activation domain; they therefore serve as repressors of transcription [6,12]. Since these nuclear receptor proteins control many important developmental and physiological processes, it is imperative to understand their control mechanism.

Two forms of the *Drosophila ftz-f1* gene products, named α and β , have been identified. *ftz-f1* α is maternally expressed and distributed in the early embryo [1], whereas *ftz-f1* β is zygotically expressed and present in the late embryo and early pupal stages [13]. Both forms play a role in ecdysone-induced gene expression [14].

Xenopus also has two *ftz-f1* genes [3], of which the gene products, xFF1rA and xFF1rB, are probably the orthologues of

mouse liver receptor homologue (LRH)-1. In addition to the full-length cDNA, another shorter gene product with a C-terminal deletion, termed 'xFF1rA-short', was also isolated [12]. xFF1rA-short is present at a lower concentration, and lacks the transactivation function.

In humans, different FTZ-F1 homologues exist. One is hB1F, which is expressed in the hepatocyte and can bind and activate enhancer II of hepatitis B virus [15]. Various mouse FTZ-F1 homologues have also been reported. Among them, steroidogenic factor 1 (SF1) and embryonal long-terminal-repeat binding protein (ELP) have overlapping transcription units, which share most exons, but have their own promoters and are differentially spliced [16]. SF1 is expressed mainly in the adrenal, testis, ovary, pituitary gonadotroph and some regions of the brain [17]. It plays an important role in the development of the above cell types and transactivation of a number of genes involved in steroidogenesis and other endocrine functions. ELP is found in embryonic carcinoma cells, with an as-yet-unidentified physiological function [18].

Another FTZ-F1 homologue in the mouse is termed LRH-1 [19]. LRH-1 and SF1 are very similar in the DNA-binding domain, but more divergent in the ligand-binding domain. Therefore there are at least two different *ftz-f1* genes in the mouse. LRH-1 is expressed mainly in the liver, but its function is still not clear at present. In the rat, the LRH-1 orthologue is termed FTF (fetoprotein transcription factor), a transcription factor for the α 1-fetoprotein gene [20]. Rat FTF is expressed mainly in the gut endodermal cells, including liver and pancreas [20]. On the basis of the expression pattern, it may be inferred that FTF functions during differentiation of gut endodermal cells.

Abbreviations used: FTZ-F1, Fushi-tarazu Factor-1; *ff1*, zebrafish gene homologue of the FTZ-F1 family of nuclear receptors (the corresponding protein product is FF1); FTF, fetoprotein transcription factor; LRH, liver receptor homologue; ELP, embryonal long terminal repeat-binding protein; SF1, steroidogenic factor 1; 5'-RACE, 5'-rapid amplification of cDNA ends; RT, reverse transcriptase; CAT, chloroamphenicol acetyltransferase; LG22, linkage group 22.

¹ To whom correspondence should be addressed (e-mail mbchung@sinica.edu.tw).

In zebrafish, we have cloned an *ftz-f1* homologue named *ffl* [6], or NR5A2, according to the new nomenclature [21] (the protein products are designated as FF1 proteins). In the present report we describe our characterization of the zebrafish *ffl* gene. We found that the *ffl* gene is differentially spliced, resulting in an active A form and an inactive B form. In addition, it contains two separate promoters, resulting in differential utilization of exons at the 5'-region. Although both promoters are used in most of the tissues examined, the downstream promoter is preferentially used during embryonic development. *In situ* hybridization experiments showed that the *ffl* transcript was found in the brain, spinal cord, mandibular arch and digestive organs. This expression pattern implies the functional importance of *ffl* in the development of digestive organs.

MATERIALS AND METHODS

Isolation of genomic clones and DNA sequencing

Genomic clones containing the *ffl* gene were isolated from a zebrafish genomic DNA library using FF1 cDNA fragments as probes. Four clones (nos. 36, 48, 55 and 37) were further analysed. The exons and introns in these clones were identified by either PCR amplification of the λ -phage lysate with exon-specific primers or hybridization with specific end-labelled oligonucleotides. The fragments spanning exon-intron boundaries were either sequenced directly from λ -DNA or subcloned into pCRII vector (Invitrogen, San Diego, CA, U.S.A.), followed by sequencing. Introns 1, 2 and 3 were fully sequenced.

Figure 1 Organization and multiple transcripts of the zebrafish *ffl* gene

(A) Structure and alternative splicing patterns of the *ffl* gene. Exons are shown by boxes and introns by intervening lines. Horizontal lines above the gene structure indicate the extent of phage genomic clones. Alternative splicing events at the 5'- and 3'-ends of the gene are shown below the gene structure. Filled-in boxes represent exons common to all transcripts. Hatched and stippled boxes are exons unique to one form of the transcripts. The antisense primers 57 and 62, used in primer extensions, are shown by semi-arrows. (B) Diagram of four transcripts encoded by the *ffl* gene. The functional domains of the FF1 protein are shown above the exons. The DNA-binding domain (DBD), regions II and III and AF-2 are indicated by the filled-in boxes. Numbers inside boxes represent exon numbers, and the locations of primers used in the RT-PCR analysis are also shown (by semi-arrows). The translated and untranslated regions are shown by wide and narrow rectangular boxes respectively.

Primer extension and 5'-rapid amplification of cDNA ends (5'-RACE)

Antisense oligonucleotides, #57 (AGTCTTTGCAGTCAC GTTGTCTT) and #62 (GTGGTCTTTGCAGTGACTATG-ATG), were used for mapping the transcription-start sites on exon 1 and exon 2 respectively. Oligonucleotides were end-labelled with [32 P]ATP, and annealed to 50 μ g of total RNA at 50 °C overnight before extension reactions, as described previously [22]. The sequencing of the genomic fragment using the same primers was performed in parallel, and the reaction products were resolved by sequencing gels. Transcription-start sites were confirmed by 5'-RACE [23], followed by cloning and sequencing of the amplified fragments.

Reverse transcriptase (RT)-PCR and Southern-blot hybridization

RT-PCR and Southern-blot hybridization procedures were carried out as described previously [24]. The locations of the primer pairs 44/47, 24/47, 4F1/7R1 and 4F1/78 for the amplification of exon 1, exon 2, form A and form B cDNA respectively, are shown in Figure 1. Their sequences are as follows: 44, AGGACAACGGACTG CAAAGACT; 24, GGACGGCGTT-TTTGAGAATCTGTGGTG; 47, CATGCTGACTGGTG AATGTACGA; 4F1, AATGTGCAAGATGGCTGACCAG; 78, GATGCCAGACTAG TCACGTCT; and 7R1, CACATC-ACGTAGTCCAGCAGTG.

Computer analysis

DNA sequences were analysed with the Wisconsin Sequence Analysis Package of GCG. The BLAST program was used to compare the zebrafish FF1 amino acid sequences with GenBank[®] and EMBL databases [25]. Multiple sequence alignments and phylogenetic trees were generated with the UPGMA method in SeqWeb (<http://www.nhri.org.tw/b5/mis/seqweb.htm>). Both TESS and MatInspector programs in the BCM launcher were used to search for potential transcription-factor binding sites [26,27]. Photographs of embryos and sections were scanned and processed using the Adobe Photoshop program.

Whole-mount *in situ* hybridization and sectioning

Adult zebrafish and embryos were reared at 28.5 °C, as described previously [28]. Sense and antisense RNA probes were synthesized by *in vitro* transcription from linearized plasmids containing FF1 cDNA fragments using T7 RNA polymerase and digoxigenin/RNA-labelling mix. Whole-mount *in situ* hybridization was carried out as described previously [28]. After *in situ* hybridization, embryos were washed in 0.1 % (v/v) Tween 20, stored in 4 % (w/v) paraformaldehyde, and photographed in 80 % (v/v) glycerol. Embryos were washed in PBS and embedded in 1.2 % agarose/5 % sucrose (w/v) prior to cryosectioning, as described previously [28]. Frozen sections were stained with Nuclear Fast Red.

Analysis of the activation function of FF1

The full-length FF1A protein (amino acids 1–516) and subsequent deletion fragments were obtained by PCR based on the original FF1 cDNA plasmid [6]. Primers used for the C-terminal deletion constructs were as follows: AAGAGAATTCTCAG-GGCACGTCTCCGTTTCAGGTGT(1–501); AGTGAATTCCG-CCCGGATCAAGGCTTTCT (1–155). The primer used to obtain the N-terminal deletion clones (43–516 and 43–434) was GACGGATCCATGGATGAGATGTGTCC (Δ 1–43). Deletion

clone 1–434 was equivalent to the FF1B ORF (where ORF is open reading frame). Restriction enzyme sites at both ends were introduced, so that the resulting DNA fragments could be cloned into the pcDNA3 expression vector (Invitrogen), and stop codon sequences were located in each lower primer for PCR. HeLa cells were chosen for the transient transfection analysis, and consensus gonadotrope-specific element (GSE)-TATA-CAT (where CAT is chloramphenicol acetyltransferase) containing the FF1 recognition site was used as a reporter [6]. Expression plasmid (1 µg), 5–10 µg of CAT reporter plasmid and 1 µg of pCMVβ-LacZ plasmid (as an internal control) were included in each transfection (60-mm dish). CAT and β-galactosidase activities were determined as described previously [29].

RESULTS

Genomic organization and multiple forms of the zebrafish *ff1* gene

We first isolated zebrafish *ff1* A- and B-form cDNAs, which are identical in sequence in most regions, but have distinct sequences at their 3'-ends [6]. To characterize the genomic structure and splicing pattern of the *ff1* gene, we screened a zebrafish genomic library with FF1A cDNA fragments. Four clones (nos. 36, 37, 48 and 55) were isolated, which mutually covered the entire *ff1* gene (Figure 1A). The zebrafish *ff1* gene consists of eight exons interrupted by seven introns. All exon–intron junctions follow the 'GT...7AG' rule for the splice donor and acceptor sites (Table 1). The A- and B-form mRNAs result from alternative splicing at exon 7. In form A, exon 7a is spliced to exon 8. The form B mRNA retains exon 7b, so that the mature mRNA is terminated at exon 7b. In addition to the 3' mRNA variants, RT-PCR and cDNA sequence analysis revealed the presence of two leader exons. Transcription starting at exon 2 is spliced to exon 3, whereas transcription starting at exon 1 skips exon 2 and is spliced directly to exon 3.

Therefore alternative promoter usage and splicing of the zebrafish *ff1* gene give rise to four mRNA variants, designated as IA, IB, IIA and IIB respectively (Figure 1B). Exons 3, 4, 5, 6 and

7a are shared by all of these mRNA variants, whereas exons 1, 2, 8 and 7b are unique to each form. Exons 1 and 2 encode the A/B domain in comparison with other nuclear receptor family members (Figure 1B). Exons 3, 4, and part of exon 5 encode the C domain, which contains two zinc-finger motifs and the FTZ-F1 box. Most of exon 5 encodes domain D (hinge region). The 3' region of exon 5, together with exons 6, 7a and 8 encode the E domain (ligand-binding domain), which contains the AF-2 core. Exon 7b encodes only two amino acids before it runs into a stop codon. Consequently, exon 1 and exon 2 transcripts produce distinct FF1 proteins with different A/B domain sequences. The A-form transcripts produce full-length FF1, whereas the B-form transcripts produce C-terminal-truncated FF1 lacking AF-2.

The transcription-start sites for both leader exons in the liver of adult zebrafish were mapped further by primer extension and 5'-RACE. One major start site in exon 1 was found approx. 222 bp upstream from the 3'-end of exon 1 (Figures 2A and 2B). Although there were two minor bands above and below the major primer-extension product (Figure 2B), they could not be confirmed by 5'-RACE. The start site of exon 2 was mapped to 351 nt upstream from the 3'-end of exon 2 (Figures 2A and 2B). The same transcription-start sites were also identified in brain, ovary and testis RNA (results not shown). Scanning the 5'-flanking sequences for putative transcription-factor binding sites, we found sequences flanking exon 1 that will bind XFD2 (*Xenopus* fork-head domain factor 2), SN (Snail), MyoD, GATA1 and SOX5, and sequences flanking exon 2 that will bind BRN2 (Brain 2), hepatic nuclear factor-3 (HNF3) and S8 (homeo-protein S8) (Figure 2A).

Properties of different forms of the *ff1* gene products

We have shown that alternative promoter usage and splicing generate FF1 forms with different A/B domains and C-termini. To determine whether different A/B domains and C-termini could lead to differences in the transcriptional activities of various forms of FF1, a series of N-terminal and C-terminal

Table 1 Nucleotide sequences of exon/intron boundaries of the zebrafish *ff1* gene

Exon sequences are shown in uppercase letters and intron sequences in lowercase letters.

Exon	Size (bp)		Exon/intron junctions	
			Donor site	Acceptor site
1	222	1/3	TGCAAAGACTTGGgtaagt.....	tcttatgttttttagCCCCACAGTTTAAA
2	351	2/3	GGAAACATGCAAGgtaagc.....	tcttatgttttttagCCCCACAGTTTAAA
3	119	3/4	TGAGAGCTGTAAGgtgtgt.....	acctcactctgcagGGCTTCTTCAAGCG
4	142	4/5	ATGAAACTGGAGGgtaaga.....	tgtccttctataagCTGTGAGAGCCGAT
5	695	5/6	CCGGGAACATAAGgtatgt.....	cactgtccccctcagGTTGACGACCAAAT
6	120	6/7	CACGGGGCAACAGgtgagt.....	tgcccttctcttcagGTGGATTACGCTCT
7a 7b	148 748	7a/8	CTGTTTCAGCTTGGgtgagt.....	tctgttttctcagACGTGAAGAACCTG
8	1697			

A Exon 1 and its upstream sequence

-520 GCCTACATTC ATCCCATTTT AATAAACGAA TACATAAAAA ATTAGATAGG
 -470 CCTATGTGCT TTTTGGACTC TGTATACAAA CATGAAAAGT ATAAACATAT
 -420 ACTTTGGAGG TGTGTTTACG TGTATTTATT TCCGTCTGTA AGGTGTGACT
 -370 TCATCTTGAT CCCGCACTGA ACACGCGTGT ATATCCTGAT CCTCACACGT
 -320 CCAATAAAAC CGTCGTGTT TGGCAGAAAC TAGTTCAAAG TTTGCGGCC
 -270 ACTATCAACC TGTTGTGTTT TTTTGATCTC GCATTTTACC ACCTGCTCTA
 -220 TTTCCATTC A GTCATTGCCG TGGACGGTCT TGTCCGCCGG TTAATCATT
 -170 CTTCAGTACT ATCTGTCCCG TCCGGCTTAT CTGTCCGTGT GTCACCTTAT
 -120 AACTGGCTCT TATCAATCTT GTCCAGCTCC ACGTGATCTG GGCAAACTA
 -70 TCAACCTCAG AAGTTTGA TATTGTTGAG GAGGTGAGAT ATGACTGCAG
 -20 GTTTAGTGCA GGTTCATGGC AGGGGGGATG AGGGCTGAGA AAGTCAAGAG
 31 ATGTCAGGCG GATGGTTTGG GGCAAAAAA CACCTCAATT TTCTATTCAA
 81 ACGACGGTGA AAGATTGGGA AAGAATGGAC GCGTTTGTG GGATATAACT
 131 AGCCACTCAG GTGGATCACC TGGTGTATTT CTGCACTAAA ATGTCCAGTA
 181 GTTTTGAGAG TGAATTAGGA CAACGTGACT GCAAAGACTT GG

Exon 2 and its upstream sequence

-480 ATTAACATGA ATAATGCATC TCCTAATTTA AATAACTATG TTTTATGTAG
 -430 CTATTTCCAC AAATTAAGT TTTAAAAATC CTTTAGCATC TTGGTGCAAA
 -380 TAAGTAAGTC GCAAAATATAT AAAAAACGCA ATTTTATGAG ATTATTTTATG
 -330 ATCTTATATC AAATTTTATT TAGCCTGATT TTCGTCTTGC CTCCAATTCC
 -280 CCCTTTCTCT ATATTTTATTA ATAGCTACAT ACAACATTTT TAACCTCCG
 -230 GTAGGCTAAA TCACCACATA CCTCCCGGTG CTCCAGTTAA CGGCGCGCGC
 -180 TGTCTCGTAA TGTTCAATTA GTTTAGTTTG TCTGTGTTTG CCGCTGCAAC
 -130 TGAGGCGCTT ATTCTCGGGC CCGTTTCTGA TTGGACAAAC GCGTGACGCC
 -80 AGCGCACTGA TCCGAACGCT GATTGGTCTGA CGCGGTGTCA GTGCGGCGTA
 -30 AACAAAGCAG CGAGGAGCCG TTTGAACTTC ACAGATAGAG CGACACGGAG
 21 CGCGCGCGGA GGAGGGGCTC GTGCCAAGTT CCTGACAGTT CGCGCACTCC
 71 GGTTTAAAGA CCAGCATCAT AGTCACTGCA AAGACCACCC GAATGAACTT
 121 ACACAGCGTG TGAAGAACTA AAAGAAAAA GAAGGGTAGT CGAAGTCAAG
 171 CCGAGGATTT GTTTCGGACG TTTGGACGCG GTTTTGTAGA ATCTGTGGTG
 221 GCGCGCAGGC GGAGCGGCGC GCGCTCATCA TCGGCTCAGC AGAGCTGTCA
 271 TGCTGCCTAA AGTCGAGTCA GAATATCTGG GACTCGCTCG ATCGCATGGG
 321 GAACAGGGGC ATATGCCTGG AAACATGCAA G

Figure 2 Zebrafish *ff1* gene has two promoters

(A) Sequences of the first and second exon and their respective 5'-upstream regions. Transcription-start sites and the translation initiation codon are indicated by horizontal arrows and M respectively. Potential transcription-factor binding sites located upstream of the start sites are indicated by lines above the recognition sequence. (B) Mapping of transcription-start sites on exon 1 and exon 2 by primer extension. A 32 P-labelled antisense oligonucleotide was hybridized to liver total RNA and extended with RT, as described in the Materials and methods section. The same primer was used in the dideoxynucleotide-sequencing reactions on a plasmid containing the exon 1 or exon 2 sequences. The primer extension product and the sequencing reactions were separated on a sequencing gel, and detected by autoradiography. The asterisk denotes the transcription-start nucleotide in the lower strand, and the arrows highlight the primer-extension products.

Figure 3 Analysis of FF1 domain function for transcriptional activation

The structure of the effector plasmids are shown on the left (Figure is not to scale), with the numbers indicating the beginning and end of the effector molecules. The effect of deletion clones were examined with a co-transfection study in HeLa cells. The reporter construct is a *sGTHIIβ* gene — 39 basal promoter linked to the *CAT* gene, with a consensus GSE sequence 5' upstream (consGSE-TATA-CAT). Basal value represents the activity of the consGSE-TATA promoter without transcription-factor co-transfection. All histograms represent the means $\pm$ S.E.M. of at least two independent experiments (in duplicate or triplicate).

Figure 4 A phylogenetic tree of FTZ-F1-related proteins

The zebrafish FF1 protein falls into the group of FTF/LRH-1 proteins. The DDBJ/EMBL/GenBank[®] accession numbers of the proteins included in the tree are: dFTZ-F1 (M63711), dFTZ-F1 (M98937), bAd4BP/SF-1 (D13569), hAd4BP/SF-1 (U76388), pSF-1 (U84399), mELP3 (AB000490), cAd4BP/SF-1 (AB002404), mdFTZ-F1 (AB016834), mLRH-1 (M81385), rFTF (U47280), hFTF (U93553), hFTZ-F1 (AB019246), hB1F (U80251), cFTF/LRH-1 (AB002403), xFF1rA (U05001), xFF1rB (U05003), zFF1 (AF014926), tFZR1 (AB006153). d, *Drosophila*; b, bovine; h, human; p, pig; m, mouse; c, chicken; md, medaka; r, rat; x, *Xenopus*; z, zebrafish; t, trout.

deletion constructs were tested (Figure 3). In co-transfection experiments, the removal of the AF-2 region resulted in a significant loss of the transcriptional activity of FF1A (as seen in clone 1–501). Proteins with clone 1–434 (equivalent to zFF1B) and 1–155 yielded a transactivation potential similar to that of clone 1–501. These data suggest that AF-2 is critical to the activation function of FF1A, and FF1B is transcriptionally inactive. N-terminal truncation of the A/B domain of FF1A (clone 43–516) did not change the transactivity of FF1A. The activity of FF1A was drastically lowered only when the AF-2

Figure 5 Tissue and developmental expression patterns for zebrafish FF1 mRNA variants

RT-PCR was performed using primers specific for different isoforms of the *ff1* mRNA, followed by Southern-blot hybridization. Actin amplification was used as a reference for the amount of total RNA input in the analysis. (A) Expression of FF1 mRNA variants in various adult tissues. B, brain; H, heart; L, liver; I, intestine; O, ovary; T, testis; M, muscle; E, eye. (B) Developmental expression of FF1 exon II, A form, and B form mRNA. C, cleavage; B, blastula; G, gastrula; S, segmentation; P, pharyngula; H1, H4, H11 and H28, 1, 4, 11 and 28 days after hatching respectively; A, adult.

and A/B domains were simultaneously deleted (clone 43–434). Therefore the A/B domain of FF1 is not directly involved in the transactivation of its target genes.

Figure 6 Expression of *ff1* in zebrafish larvae

Views are anterior right and posterior left. (A) Lateral view of zebrafish 96 h post-fertilization after whole-mount *in situ* hybridization. (B and C) Dorsal views of the horizontal sections of whole-mount, *in situ*-hybridized larva. Expression of *ff1* was seen in the mandibular arch (M), hypothalamic nucleus (H), spinal cord (S), pancreas (P), liver (L) and intestine (I). Y: yolk. Arrows indicate the planes of the horizontal sections shown in (B) and (C).

Zebrafish FF1 is a member of the vertebrate FTF/LRH-1 family

To understand the link between zebrafish FF1 and other FTZ-F1-related proteins, multiple-sequence alignments and phylogenetic analysis were performed. Compared with other FTZ-F1-related proteins, zebrafish FF1 shares higher identity (78–84%) with the frog (xFF1rA, xFF1rB), chicken, murine and human FTF/LRH-1 proteins than with the chicken, murine, bovine and human Ad4BP/SF-1 (58–66%). Phylogenetic tree analysis shows that the vertebrate FTZ-F1-related proteins can be divided into two subgroups (the Ad4BP/SF-1 and FTF/LRH-1 subgroups), where zebrafish FF1 protein falls into the subgroup of FTF/LRH-1 (Figure 4). Therefore it should be assigned with the name NR5A2, according to the new nomenclature [21].

Spatial and temporal expression of FF1 mRNA variants

To examine the expression profiles of FF1 mRNA variants, RT-PCR was performed to detect their expression in various adult tissues and developmental stages (Figure 5). *ff1* was expressed in many tissues, including brain, liver, intestine, muscle, eye and gonads. All four mRNA variants were similarly expressed in these tissues with some quantitative variation, with the exception

that exon 1 transcripts appeared to be missing in the brain. During development, exon 2 and form A transcripts were first detected in the segmentation stage and expression continued through the pharyngula, larval and adult stages (Figure 5B). Form B transcripts were expressed at low levels in the embryonic and larval stages, whereas exon 1 transcripts could only be detected after amplification with more cycles (results not shown).

Whole-mount *in situ* hybridization was performed to probe further the site of *ff1* expression during early development. Strong signals were detected only when probes specific for exon 2, form A or common to all regions were used for hybridization. No signal was observed in embryos hybridized to RNA probes specific for either exon 1 or the B form transcripts (results not shown). This result was consistent with the results of RT-PCR, in which transcript IIA was the major form expressed in early zebrafish development. Whole-mount *in situ* hybridization revealed *ff1* transcripts in the developing hypothalamus, mandibular arch, spinal cord and gut derivatives, such as intestine, pancreas and liver (Figure 6A). Sectioning of whole-mount embryos confirmed further that *ff1*-expressing cells were located in the hypothalamus, neurons of spinal cord, mandibular arch, liver and exocrine cells of pancreas (Figures 6B and 6C). Zebrafish GATA-5 [30], used as a marker for liver and pancreatic exocrine

cells, showed a similar expression pattern in the gut derivatives (results not shown).

DISCUSSION

In this report, we analysed the zebrafish *ftz-fl* (*ffl*) gene structure and expression. We showed that the presence of two promoters and alternative splicing of the *ffl* gene results in four transcripts encoding proteins with different N- and C-terminal sequences. Alternative splicing at the 3'-end generates transcription factors with opposite functions, whereas differential promoter utilization does not produce an apparent effect on protein function in our assay system. These transcripts are similarly expressed in many adult tissues. During embryogenesis, however, there exists only a major transcript, IIA. IIA transcripts are detected mainly in brain and gut derivatives during embryonic development, thus implying that they function in the development of these cell lineages.

The zebrafish *ffl* gene belongs to the FTF/LRH-1 subgroup in the FTZ-F1 family of the nuclear receptor superfamily. This conclusion is on the basis of two lines of evidence, i.e. sequence comparison studies and the pattern of gene expression. Our phylogenetic analysis studies have shown that the vertebrate FTZ-F1 family can be divided into the Ad4BP/SF-1 and FTF/LRH-1 subgroups (Figure 4). These two groups have probably arisen by gene duplication specific to the vertebrate lineage [3]. Zebrafish FF1 falls into the subgroup of vertebrate FTF/LRH-1 because its sequence in several regions, including the D box in the second zinc finger, region II, region III and helical segments H9 and H10 in the C-terminus, is more similar to that of FTF/LRH-1 than that of Ad4BP/SF-1.

The zebrafish *ffl* gene is located on linkage group 22 (LG22) [31]. The only other known gene in this linkage group is the zebrafish *SCL* gene [31]. Another human FTF/LRH-1 subgroup member, *B1F* [15], and the human *SCL* gene [32] are both located on chromosome 1. It is commonly observed that chromosome segments are conserved across species, i.e. when genes on one chromosome segment are linked, they are also often linked on the chromosome from a different species [31]. Although it may be coincidental, the linkage of these two genes on both zebrafish and human chromosomes could indicate similarity between zebrafish LG22 and human chromosome 1. To date, very few genes have been mapped to zebrafish LG22. When more genes are mapped, the possibility of conservation between LG22 and human chromosome 1 can be further evaluated.

Zebrafish *ffl* is highly expressed in many tissues, but mainly in gut derivatives such as liver and pancreas, as shown by RT-PCR and *in situ* hybridization. This strong hepatic or pancreatic expression is also found for genes of the FTF/LRH-1 subgroup [3,15,20,33], but not for the *Ad4BP/SF-1* genes, which are expressed primarily in adrenal glands and gonads [4,16,34]. Therefore both phylogenetic and expression pattern analyses lead to the assignment of zebrafish *ffl* to the FTF/LRH-1 subgroup.

Since most vertebrates have *ftz-fl* genes belonging to both *Ad4BP/SF-1* and FTF/LRH-1 subgroups, zebrafish should also have its members in both subgroups. Because zebrafish chromosome segments are often duplicated [31], it is possible that the zebrafish *ftz-fl* genes are also duplicated; therefore more than two *ftz-fl* genes could be present. Consistent with our prediction, many *ftz-fl* cDNA fragments have been isolated in zebrafish (D. Liu and C. L. Hew, unpublished work).

Zebrafish FF1 mRNA is expressed in the developing gut, liver and pancreas. It implies that *ffl* may function in these gut

derivatives. Supporting this hypothesis, the FF1 homologue, human B1F, functions as a hepatocyte transcription factor to activate hepatitis B virus enhancer [15]. Rat FTF has been shown to activate α 1-fetoprotein transcription in liver cells [20]. Furthermore, the locations of zebrafish *ffl* expression in the gut derivatives are the same as those of zebrafish GATA-5 expression [30]; W.-w. Lin, unpublished work). Thus the co-localization with zebrafish GATA-5 suggests that the function of FF1 might be related to the action of GATA-5 in the development of gut derivatives. The mouse GATA-5 homologue, GATA-6, is required for the differentiation of visceral endoderm [35]. Consistent with the notion of expression in gut derivatives, we found sequences in the 5'-flanking region of the *ffl* gene that will bind the gut-enriched transcription factors HNF-3 (hepatic nuclear factor-3) and GKLF (gut-enriched Kruppel-like factor). In addition, consensus sequences that will bind brain transcription factor BRN2, gonad factor SOX5, and factors related to early embryogenesis, such as MyoD, XFD2 and Snail, are also present in this region. Whether these sites function in the transcription of the gene remains to be elucidated.

Apart from the gut derivatives, we also detected *ffl* mRNA in the mandibular arch, and neurons in the spinal cord and hypothalamus (Figure 6). The significance of *ffl* expression in these sites is not yet clear. One possible link is that the formation of hypothalamic nuclei is known to require SF-1 [36]. It remains to be elucidated whether *ffl* plays a role similar to that of SF-1 in the development of the hypothalamus. It is also possible that other uncharacterized Ad4BP/SF-1 family members are important for the development of the hypothalamus.

To date, *ffl* is the only member of the FTF/LRH-1 subgroup for which the gene structure has been characterized. Its structure is very similar to the genes in the Ad4BP/SF-1 subgroup [5,16]. This suggests that the gene structure in the *ftz-fl* family is conserved. Multiple promoter usage and alternative splicing are common features shared by members of the *ftz-fl* family [5,16], which generate protein diversity and allow for regulation of gene expression at multiple levels. All four forms of FF1 mRNA variants are expressed similarly in the gut derivatives and gonads, except in the brain, where only exon 2 transcripts are detected (Figure 5). We also detected predominant expression of exon 2 and A form transcripts during early development. This suggests that promoter 2 is probably used preferentially, and the full-length A form receptor rather than the truncated B form prevails to exert its activation function in early development.

Exon 1 and 2 transcripts encode FF1 proteins with different A/B domains (Figure 1). The A/B domain usually contains the AF-1 function important for ligand-independent transactivation through interaction with other co-factors [8,37–39]. We show that the A/B domain does not contribute towards transcriptional activity in HeLa and COS-1 cells, both of which do not normally express FF1 (Figure 3; also D. Liu, unpublished work). However, we cannot exclude the possibility that the A/B domain may play a role in some tissues where appropriate cofactors are present. Since the brain expresses only exon 2 transcripts, it might imply the existence of functions specific to the brain.

Compared with the mouse *ftz-fl* gene, there is a difference in the 3' alternative splice site. Zebrafish *ffl* is alternatively spliced at exon 7, whereas mouse *ftz-fl* is at exon 6 [16]. This results in proteins truncated at different positions. Alternative splicing at exon 7 generates FF1A and FF1B, which function differently in their transactivation potential. FF1B retains regions II and III important for DNA binding [12], but loses the ability to transactivate. Therefore it could be a transcriptional repressor by competing for DNA binding [6]. Other 3'-splicing variants, like mELP1 and xFF1rA-short, lack region III and therefore bind

DNA at a reduced efficiency. High amounts of these short forms are required to achieve the dominant negative effect *in vitro* [12].

We thank Chi-Yao Chang, Sheng-Ping Hwang, Jye-I Wang and John Postlethwait for the zebrafish genomic DNA library, fish maintenance and helpful discussions. This work was supported by grants NSC87-2311-B-001-032-B24 and NSC86-2745-B-001-001-B24 from the National Science Council, and by Academia Sinica, Republic of China to B.-c.C. and funding support from the Medical Research Council of Canada to C. L. H. D. L. is a recipient of a Research Training Fellowship from the Hospital for Sick Children in Canada.

REFERENCES

- Lavorgna, G., Ueda, H., Clos, J. and Wu, C. (1991) *Science* **252**, 848–851
- Ueda, H., Sun, G. C., Murata, T. and Hirose, S. (1992) *Mol. Cell. Biol.* **12**, 5667–5672
- Ellinger-Ziegelbauer, H., Hihi, A. K., Laudet, V., Keller, H., Wahli, W. and Dreyer, C. (1994) *Mol. Cell. Biol.* **14**, 2786–2797
- Ikeda, Y., Lala, D. S., Luo, X., Kim, E., Moisan, M.-R. and Parker, K. L. (1993) *Mol. Endocrinol.* **7**, 852–860
- Wong, M., Ramayya, M. S., Chrousos, G. P., Driggers, P. H. and Parker, K. L. (1996) *J. Mol. Endocrinol.* **17**, 139–147
- Liu, D., Le Dren, Y., Ekker, M., Xiong, F. and Hew, C. L. (1997) *Mol. Endocrinol.* **11**, 877–890
- Lala, D. S., Rice, D. A. and Parker, K. L. (1992) *Mol. Endocrinol.* **6**, 1249–1258
- Tsai, M.-J. and O'Malley, B. W. (1994) *Annu. Rev. Biochem.* **63**, 451–486
- Mangelsdorf, D. J. and Evans, R. M. (1995) *Cell* **83**, 841–850
- Enmark, E. and Gustafsson, J. A. (1996) *Mol. Endocrinol.* **10**, 1293–1307
- Ohno, C. K., Ueda, H. and Petkovich, M. (1994) *Mol. Cell. Biol.* **14**, 3166–3175
- Ellinger-Ziegelbauer, H., Glaser, B. and Dreyer, C. (1995) *Mol. Endocrinol.* **9**, 872–886
- Ohno, C. K. and Petkovich, M. (1993) *Mech. Dev.* **40**, 13–24
- White, K. P., Hurban, P., Watanabe, T. and Hogness, D. S. (1997) *Science* **276**, 114–117
- Li, M., Xie, Y. H., Kong, Y. Y., Wu, X., Zhu, L. and Wang, Y. (1998) *J. Biol. Chem.* **273**, 29022–29031
- Ninomiya, Y., Okada, M., Kotomura, N., Suzuki, K., Tsukiya, T. and Niwa, O. (1995) *J. Biochem. (Tokyo)* **118**, 380–389
- Parker, K. L. and Schimmer, B. P. (1993) *Trends Endocrinol. Metab.* **4**, 46–50
- Tsukiya, T., Ueda, H., Hirose, S. and Niwa, O. (1992) *Mol. Cell. Biol.* **12**, 1286–1291
- Tugwood, J. D., Issemann, I. and Green, S. (1991) GenBank® Nucleotide Sequence Database accession no. M81385
- Galarneau, L., Pare, J. F., Allard, D., Hamel, D., Levesque, L., Tugwood, J. D., Green, S. and Belanger, L. (1996) *Mol. Cell. Biol.* **16**, 3853–3865
- Nuclear Receptors Committee (1999) *Cell* **97**, 1–20
- Guo, I.-C., Huang, C. and Chung, B.-c. (1993) *DNA Cell Biol.* **12**, 849–860
- Frohman, M. A., Dush, M. K. and Martin, G. R. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 8998–9002
- Lai, W. W.-K., Hsiao, P.-H., Guiguen, Y. and Chung, B.-c. (1998) *Endocr. Res.* **24**, 927–931
- Altschul, S. F. and Lipman, D. J. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 5509–5513
- Quandt, K., Frech, K., Karas, H., Wingender, E. and Werner, T. (1995) *Nucleic Acids Res.* **23**, 4878–4884
- Smith, R. F., Wiese, B. A., Wojzynski, M. K., Davison, D. B. and Worley, K. C. (1996) *Genome Res.* **6**, 454–462
- Westerfield, M. (1995) *The Zebrafish Book: A Guide for the Laboratory Use of Zebrafish (*Danio rerio*)*, University of Oregon Press, Eugene
- Drean, Y. L., Liu, D., Wong, A. O., Xiong, F. and Hew, C. L. (1996) *Mol. Endocrinol.* **10**, 217–229
- Pack, M., Solnica-Krezel, L., Malicki, J., Neuhauss, S. C., Schier, A. F., Stemple, D. L., Driever, W. and Fishman, M. C. (1996) *Development* **123**, 321–328
- Postlethwait, J. H., Yan, Y. L., Gates, M. A., Horne, S., Amores, A., Brownlie, A., Donovan, A., Egan, E. S., Force, A., Gong, Z. et al. (1998) *Nat. Genet.* **18**, 345–349
- Begley, C. G., Aplan, P. D., Davey, M. P., Nakahara, K., Tchorz, K., Kurtzberg, J., Hershfield, M. S., Haynes, B. F., Cohen, D. I., Waldmann, T. A. et al. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**, 2031–2035
- Kudo, T. and Sutou, S. (1997) *Gene* **197**, 261–268
- Parker, K. L. and Schimmer, B. P. (1997) *Endocr. Rev.* **18**, 361–377
- Morrissey, E. E., Tang, Z., Sigrist, K., Lu, M. M., Jiang, F., Ip, H. S. and Parmacek, M. S. (1998) *Genes Dev.* **12**, 3579–3590
- Ikeda, Y., Luo, X., Abbud, R., Nilson, J. H. and Parker, K. L. (1995) *Mol. Endocrinol.* **9**, 478–486
- Hadzic, E., Desai-Yajnik, V., Helmer, E., Guo, S., Wu, S., Koudinova, N., Casanova, J., Raaka, B. M. and Samuels, H. H. (1995) *Mol. Cell. Biol.* **15**, 4507–4517
- Tora, L., White, J., Brou, C., Tasset, D., Webster, N., Scheer, E. and Chambon, P. (1989) *Cell* **59**, 477–487
- Onate, S. A., Boonyaratankornkit, V., Spencer, T. E., Tsai, S. Y., Tsai, M. J., Edwards, D. P. and O'Malley, B. W. (1998) *J. Biol. Chem.* **273**, 12101–12108

Received 15 October 1999/11 January 2000; accepted 2 March 2000