ELSEVIER

Biochimica et Biophysica Acta 1395 (1998) 50–56

BIOCHIMICA ET BIOPHYSICA ACTA
BBA

Promoter paper

Genomic organization and characterization of the promoter region of the round-spotted pufferfish (*Tetraodon fluviatilis*) *JAK1* kinase gene [1]

Jiann-Horng Leu [a], Mau-Sun Chang [b], Chen-Wen Yao [c], Chen-Kung Chou [d], Shui-Tsung Chen [a], Chang-Jen Huang [a,*]

[a] *Institute of Biological Chemistry, Academia Sinica, Taipei, Taiwan*
[b] *Department of Zoology, National Taiwan University, Taipei, Taiwan*
[c] *Graduate Institute of Life Science, and Institute of Preventive Medicine, National Defense Medical Center, Taipei, Taiwan*
[d] *Department of Medical Research, Veterans General Hospital, Taipei, Taiwan*

Received 2 May 1997; accepted 21 July 1997

Abstract

Seventeen kilobases of genomic DNA containing the promoter and the coding region of the round-spotted pufferfish *JAK1* gene was isolated and completely sequenced. This gene consists of 25 exons and 24 introns spanning about 13.5 kb, compared to >30 kb in carp *JAK1* gene. Primer extension analysis revealed one transcription initiation site which was 376 bp upstream of the translation initiation site. The sequence of the 2.9 kb region upstream of the transcription initiation site contains numerous potential binding sites for transcription factors including HNF-5, GCF, Sp1, CRE, AP2, GATA, GAGA, E2A, p53, and NF-IL6. When this region was placed upstream of the chloramphenicol acetyltransferase (CAT) reporter gene and transfected into a carp CF cell line, it could drive the synthesis of CAT enzyme three times more efficiently than could the common carp *JAK1* promoter. © 1998 Elsevier Science B.V.

Keywords: *JAK1* kinase; Genomic structure; Promoter; (Round-spotted pufferfish); (*Tetraodon fluviatilis*)

The *Janus* kinases (JAKs) family belongs to the non-receptor protein tyrosine kinases and currently consists of four members in mammalian species, i.e. JAK1, JAK2, JAK3 and TYK2 [1,2]. The JAKs are known to be involved in many cytokine signaling through the JAK–STAT pathway [3]. In addition to mammalian JAKs, two JAK homologs have been cloned and characterized in invertebrate [4] and fish [5]. In *Drosophila*, a single JAK homolog, encoded by the gene *hopscotch* (*hop*), has been identified and shows the highest degree of identity with mammalian JAK2. Moreover, a putative *Drosophila* STAT protein has also been identified. Thus, the existence of an invertebrate JAK/STAT system has been established [6]. In fish, we have cloned a 3.7 kb cDNA which encodes the carp JAK1 kinase and also isolated and characterized its genomic clone. This gene consists of 24 exons and 23 introns spanning more than 30 kb [5]. For the comparative genomic structure analysis, we reported the cloning of *JAK1* kinase gene from a round-spotted pufferfish (*Tetraodon fluviatilis*) in this report. The round-spotted pufferfish is

* Corresponding author. Fax: +886-2-788-9759; E-mail: cjhuang@ccvax.sinica.edu.tw

[1] The sequence data in this paper has been submitted to the EMBL/GenBank Data Libraries under the accession number U53213.

PII S0167-4781(97)00137-1

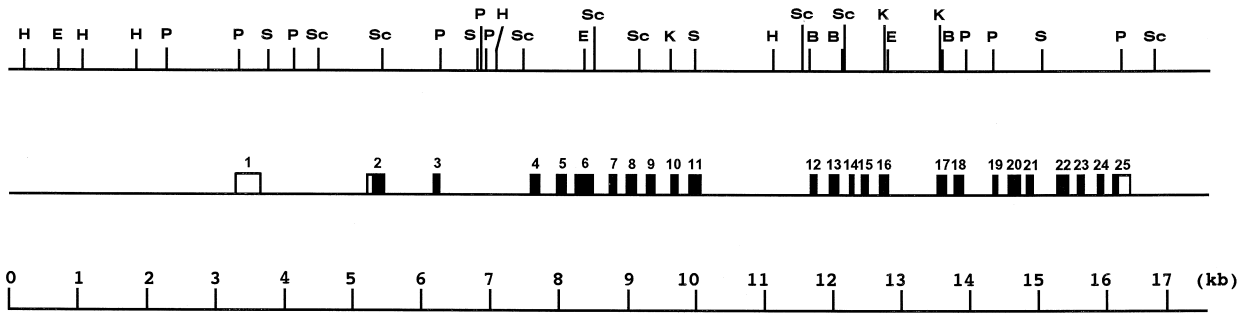

Fig. 1. Organization and physical map of the round-spotted pufferfish *JAK1* gene. Exons are indicated to scale by boxes numbered 1–25. Solid boxes indicate the round-spotted pufferfish *JAK1* coding region whereas open boxes represent the 5′- and 3′-untranslated region. Introns and the 5′- and 3′-flanking regions are indicated by the solid lines. A restriction map was indicated to contain several cleavage sites. Restriction endonuclease sites are B, *Bam*HI; E, *Eco*RI; H, *Hin*dIII; K, *Kpn*I; P, *Pst*I; S, *Sal*I; Sc, *Sac*I.

easily obtained at very cheaper prices from a local aquarium and has a genome size of 380 Mb which is the smallest of any vertebrate [7,8]. A different species of the pufferfish *Fugu rubripes* (*Fugu*) has been shown to have a compact genome with small introns, and thus is used as a model for vertebrate genome

Table 1
Exon–intron organization of the round-spotted pufferfish *JAK1* gene

exon number	exon size (bp)	3'end of the exon	5'end of the intron	intron size (bp)	3'end of the intron	5'end of the exon	amino acid interrupted
1	368	CTA AGC TCC TAG	gtgagtaaat	1646	gcgtttacag	T GTC TGA ATG	
2	212	AGG AAG TGC T	gtaagtgggg	637	ctttttccag	CA ATC TCT CCT	Ser 69 (1)
3	124	TAT CGG ATG AG	gtaagaacac	1403	cttgtgtcag	A TTT TAT TTC	Arg110 (2)
4	154	TTG TTC TAC CAG	gtgattatac	228	tttggtgcag	GGC CAG CAC	Gln161 (3)
5	152	CAT GAT ATC AG	gtaaaactaa	106	gcccttccag	T TAC AAA CGA	Ser212 (2)
6	269	GGA TAC TAT C	gtaagattcc	200	tcgattttag	GG TAT TTC AAC	Arg302 (1)
7	110	AAA CCT GAA ATG	gtgagcgaaa	106	catgttgcag	GCA CTG ATA	Met338 (3)
8	186	AAC ATG GGC ATG	gtgagctcat	117	tctttttcag	GAA TTG CAA	Met400 (3)
9	158	GGA CCC ATC AG	gtcagtactg	201	caacctcaag	C ACT GAG TAT	Ser453 (2)
10	115	GTC TGC ACT GAG	gtagacttgg	129	ctctcttcag	CTT GAC CTG	Glu491 (3)
11	199	CAG CCA AGA G	gtgagcttat	1638	tgttttgcaq	AG ATC TCT AAC	Glu558 (1)
12	110	GAC ATC GAG CAG	gtatgacagt	137	gctgttctag	GAG GAG CAC	Gln594.(3)
13	153	GAC ATC TCT TTG	gtgagaggaa	128	ttgtcggcag	GCT TTC TTT	Leu645 (3)
14	88	CAT CAG GAG A	gtaagtagat	78	tgtggaacag	AT ATC ATG GTT	Asn675 (1)
15	128	CTC AGC TAC TTG	gtgcgttggt	120	gatgttttag	GAG GAC AAG	Leu717 (3)
16	139	ACA AGG GAG G	gtgagtcaaa	710	gcttatgcag	AG TGT GTG CAT	Glu764 (1)
17	152	AAA CTT ACA GAG	gtttgaggcc	75	tatctttgag	CAG AAG GAG	Glu814 (3)
18	154	GGA GAA CAG A	gtaagttgga	450	cattcctcag	AT CCG TCC ATC	Asn866 (1)
19	89	GAC CTA GGA GAG	gtacgccctt	141	tctctcacag	GGG CAC TTT	Glu895 (3)
20	193	CAG GAA GAA G	gcaagctgta	68	gtctaaacag	GC GGT CAG GCT	Gly960 (1)
21	125	CAG ATA TGC AAG	gtaagactgg	386	tgttgtttag	GGG ATG GAA	Lys1001(3)
22	173	CCT GTT TTC TG	gtaggtcagg	93	tgtgttgcag	G TAC GCC CCC	Trp1059(2)
23	118	AGT CCC ATG ACG	gttagtggct	195	gcccgtgcag	TGT TTC CTG	Thr1098(3)
24	111	TGC CCC GAG CCT	gtaagaagct	122	tctttggcag	GTT TAT GAG	Pro1135(3)
25	278	CCCCGACCCCCCCCCGACTGACCCCAACAGTCCCGTTTGAGGGTGAATAAATCTGTGGTGTTGGATCTTCTATAAAAAAAAAAAAAAAAA					
		(the end of round-spotted pufferfish JAK1 gene)					

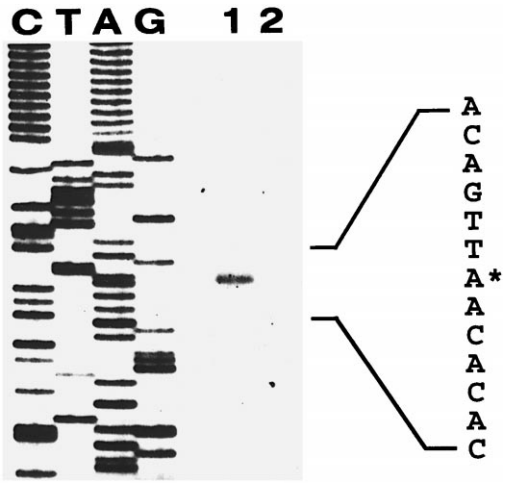

Fig. 2. Determination of the transcription initiation site of the round-spotted pufferfish *JAK1* gene. [^{32}P]-labeled primer (underlined in Fig. 3) was annealed to 10 μg yeast tRNA (lane 2, as a negative control) or 5 μg poly (A)$^+$ RNA from round-spotted pufferfish liver (lane 1) and extended with reverse transcriptase. Lanes C, T, A and G are sequencing reactions using the same primer and plasmid construct that carries the 3 kb *Hin*dIII fragment from phage clone PFJ1. The base corresponding to the transcription start site is labeled by an asterisk within the fragment of genomic DNA sequence shown to the right.

analysis [7,9]. By comparing the genomic organization and intron sizes of the round-spotted pufferfish *JAK1* gene with those of the common carp *JAK1* gene [5] in this report, our results indicate that the round-spotted pufferfish is also a good model organism for comparative vertebrate genomic structure analysis.

By using the lambda FIXII as a cloning vector (Stratagene, La Jolla, CA, USA), a round-spotted pufferfish *Tetraodon fluviatilis* liver genomic library was constructed and contained approximately 5×10^5 independent clones. The amplified library was then used to isolate 15–18 kb genomic DNA clones containing the gene that encodes round-spotted pufferfish JAK1. The carp *JAK1* cDNA [5] was labeled using a DIG DNA Labeling Kit (Boehringer Mannheim, Mannheim, Germany). Approximately 1×10^6 amplified clones were plated at a density of 5×10^4 plaque forming units/150-mm Petri dish. Hybridization and washing were carried out as previously described [5,10]. Seven positive phage clones (PFJ1 to PFJ7) were isolated. After restriction enzyme mapping, these clones were found to be the same. Therefore, only the PFJ1 clone was further characterized and sequenced. As shown in Fig. 1, the restriction map of the PFJ1 clone was constructed by digesting the phage DNA with a panel of restriction enzymes separately or in various combinations: *Bam*HI, *Eco*RI, *Hind*III, *Kpn*I, *Pst*I, *Sac*I, and *Sal*I.

A total of 17 kb of the round-spotted pufferfish *JAK1* gene was completely sequenced by conventional subcloning strategy combined with automated sequencing and deposited in the GenBank with an accession number U53213. To characterize the 5′- and 3′- end of the round-spotted pufferfish mRNA, 5′ and 3′ RACE assays were performed (data not shown). All taken together, the round-spotted pufferfish *JAK1* gene is composed of 25 exons that span about 13.5 kb of DNA. The sequences around the exon/intron boundaries were determined and shown in Table 1. All exon/intron boundaries identified conformed to the GT/AG splice donor/acceptor rule [11] except exon 20, whose 5′ donor splice-site sequence begins with GC. All coding exons were relatively small, ranging from 88 bp (exon 14) to 269 bp (exon 6). The size of introns varied considerably, ranging from 1646 bp (intron 1) to 68 bp (intron 20) with an average of 380 bp. The first exon contained the 5′-untranslated region, and the second exon contained the putative translation initiation site. The JH2 domain was located on exons 12–18, and the catalytic JH1 domain was located on exons 19–25. Exon 25 contained the last 35 amino acids as well as the 3′-untranslated region.

The open reading frame of those 24 exons encodes a protein of 1169 amino acids with a molecular mass of 132 kDa. All members of the JAK family have seven homologous domains in the molecule that have been named as JHs or JAK homology domains. JH1 is a C-terminal kinase-catalytic domain whereas JH2 is a kinase-like domain and other five JHs are present in the far N-terminal part. Amino acid sequence comparison of round-spotted pufferfish JAK1 with common carp [5], human and murine JAK1 [1,12]

Fig. 3. Nucleotide sequence of the 5′-flanking region of the round-spotted pufferfish *JAK1* gene. Potential binding sites for a variety of transcription factors are underlined. Candidate transcription start site by primer extension (see Fig. 2) is numbered as +1.

```
                    HNF-5
-2922 CGCTAAGCTCAGTGTTTGTAGCTGAAAGAATAGGTGAAGTCCTCTGTGTTTACATTGGGGTTTACACAGTATGGACGCGCTTTCTCTGCGCGTTCTCCCT
                             GCF
-2822 CCTACGTGACCGTGAACGCACCAGCGCGCCTTTCCGGTCCTCCCCGTTTAGTTGTTCAGCCTTCCCGTGTTGTTTTTAACATCTGTAACACAACATTTTT
                                    GCF                               Sp1                     GCF
-2722 TTTTTATTTTTGGACGGGCTCCGTGTACGGAGCAGCGGGGCTAATGGAGTTGTTTACATGAAGTGCGGAGGGCGGAGTAACGACGGCGAGCAGCCGCCGC
                   Sp1                         CRE
-2622 GACGTCGGCCATCGGAGGGGGCGGAGCCTGCGGGGGAATCATAACAAGTGTGACGTCACTCGGCCTTGTGGTCCACGGAGTTCTACCGGGAAGGAGTCGG

-2522 GCCCAGGGTTCCGTCTAACTCATGACAGCCTCAGCTCTGAATTCAGATTTAAAACAAGGATTAGGGCAGGGATGTCAAGATGGATATGTTTTTGGATA
      TATA box
-2422 TAAAGTTTGCCATATTTCTTATTTCTGTTTTTGCTTTCGTCTTTCTAGTATTTATTTCTGACGCTGCTACTGCTGGAACACTACAATTCCCCCTTTGTGG

-2322 AAAAAAATAGGCATATTCTTATTCTAAGTCTGAATTAACCACTGGTGGTTGAACTTTCACAGTGTCTTCAATAAACTGTTTAGGGTTGTATTTAAAATTG

-2222 CTACGTTACTCCTAATCCTCCTCCTACTACAACTAGTAGCAATAATAGCAATAAATAATTGAAAGTGTACATTTTGAATGTTTTATACCGTTTATGCTTT

-2122 TTAATTATACTTTTTTGTTTTATCTAAGTAGTTTTACTTTTTACCTTTTTTTGTAAGCTTATTTTGTATATTTTCCCCCTCAACAATACAGGGGATTGGT
          CCAAT                                                                      CCAAT
-2022 GAAAACCAATTGTTACAAGTTAATGTTTGAAATTAATCAATGAGAAATTTGTCAGCGGTGAAAAAAGCTCATCAAAATAATTTACCAATAAGAAGGAATT
                                                                     Sp1
-1922 GTTAGGGTGTCACCTTCTGATATGTGTCCTATCTTTAGTAGCTGAGCGCAACATTCATCCCTGGGCGGACGATGCCAAGGGTGGGACTGCTAGAAATGAC
        AP2                                                       GATA
-1822 GGGCTGCCCTGGTTCAGTTATCACAGTGCTGATGCTGTCCCCATCGAAGGTACAGTACTGTGATAACTGAAGGATGGCTGCCAACTTGGCACGAAGAACC
                                   GAGA
-1722 AAAACCAGCTGGTTAAAGCGAGACGTGGGAACCGAGAGAGAAGAAAAACTTGCGGCTTTGATTCAGAGAGGTGATGCTTGAATTTCACTTTTTCTTGTGT

-1622 CAATATTGTAAATGTTTAAATGTGGCTTGCTCAGTTCGAATTTTGTTATTTCATCATTATTTTTATCCACGTCCAGCGTCAAGAAATAATGACTAACTTT

-1522 GCTGAGCAGAATACACAAATGGCCTAATTCCTTCTGCTCACTAACATCTAACATCATGTCAGAGCTTGTCAATATTTAGTTCCCGGTTAATCAAAAGCAT
                                         CCAAT                      GCF            E2A
-1422 CAAAGCATAGTTGGATGAGTCACACCTCTCATTTAAAATTCCAATGATTCCTTAAAGCTTCATCCTGGAACCGGCGCCATCGTCACACAGGTGTGAAAGG
                                                                                          TATA box
-1322 TTCAGGGTCCATCAGCAGGGGGTGGTAAAGATCCATCCTGGGTGAAAAGCTCCTGCTCCATTCTAACGCGTTTCCTGTTCTCTATAAATGACTGCGTATT

-1222 TAGAGTTTATTATCATAAAATAATGCCCAGAAAGAACTGTTGGAAAAAAAAATAGATGTGTGATTTATCCGAGTTATTGTTTTCTTACTAAACAGGGAATA

-1122 ATAATAATATTGTTGATTTGCTATCAACAACTTAAGTGAACTTGAATAAGTTATATTTCTTAACTTTGCAGAACTCTTATTGGAATTTTTTTAAAATGCC
                                                                      E2A                     P53
-1022 CTTTTAGTCATACAAGTTCAGCTTTGGTTGATTCAGGGCATTTCTGTTGTGTGGACTGATCCTGCAGTTGTGGAAATCCCCACTTACTGGGGCTTGTTTT

-922  TGTCTCCATTTAACATCTGTTATTGTACTTTTCTCACACATTACACTTTGTGGTAATTTTCACATACATAACACTTTTTTCACGCACTTTTTGAACGCAA

-822  ATTTAATGTTTATATTCTTTTTTGCTGACATTTTTAACATAGAGGGTAAAGTTTTTTTCTTTTCTCGAATTTGCACTTCCTTAAATAAAAACAAAGAGTA
                   HNF-5                                                    NF-IL6          HNF-5
-722  CCTTCTTTGATGTGGTATGTTTGTTTCCTCTTTTTCAAATGGTCCACCCATTGTGCACGTCCTCTTACTTCCATTTTGTAAGTTTTAAAACAGCTGTTTG
                                NF-IL6
-622  CTAGGAGAGACTCTAGAATCGAATTAGGAATTAGGGAAGGAGAAAAGATGATCGAATACACATTAGGTATGGATTTTGCATTTATTATAAGCAACATTAA
                             NF-IL6                                                P53
-522  CCCTGTATAAATGGTTGAAAACAACATTGTGTAATCAGTACATTTTTGAGTGTGATTTGAGTAGAATACAAGACAAAAACTTGTTTCAGGGGCAGTGGTC

-422  AATGCCAAGTAAAAAGTGACTGGTGAATTGAGGAGAAAAAGCAGGACTCGTGGTTCAAGAGACAAGACAACACACACAAACACACGTGTTTTCAAATGTG

-322  TCTACAATAAAAATATACTGTTAACCCTTAAAGTTATCTGTGATTTTATAGTAAATTGTAAGTTATAGTAAATTATAGTATTCTAAAAAGAAAAGCAAAC

-222  GTGCATTTTACTAAATACACTTACTGAACGTTATATTTACATGTTCTTCCATTAAACATTTCGACTTTAAATGTTTTTCTTTGTAGAGGTTGGTGAAAAT

-122  GATAAATGCAGAATCTGTGGCACTACACAAAATCACTATGGAAAATGTATCCCTTTATTTCTTCTTGCACACACACACACACACACACACACACACACAA
                          +1
-22   AGTCATATTTCTGTCCACAGTTAACACACAGACGGCGTACATACGCGCAGAACGCGTCACTTTGATTGACGGCTGCAGCAGCCAATCAGCTGGCTTCTTT
                         (transcription start site of puffer fish JAK1 gene)
+79   TAGCCAGGACTCCTGCGTCGCTTTCGGGAAGGGGCCGATACTTCTCGGGGTTAGCAAGAACGGAGGTGGGCTGCTACACTCGGGTCTAACTCTGAAATAC
                                                 (ending of 5' RACE products).
+179  GCTTTAAAAACTACGATTTATCGAGTTCATATCATCCAGAAGAGATTTCGGATATTTACTTTGGACGAAACAAAATTGTCGATAGGCGGTGTTTGGAGTC
```

reveals that there is a higher sequence homology in JH2 domain (81.5% identity for carp; 70% for human/mouse) and in JH1 domain (70% identity for all species). The overall sequence identities between round-spotted pufferfish JAK1 and common carp JAK1 are 66.4% whereas the identities between the round-spotted pufferfish JAK1 and human/mouse JAK1 are approximately 57%.

The transcription start site was determined by primer extension analysis using poly $(A)^+$ RNA from the round-spotted pufferfish liver. We used a 24-mer oligonucleotide labeled with ^{32}P at the 5′-end. The exact position of the extended product was determined by aligning the sequencing ladder obtained with the same primer. One major extended product was revealed and corresponded to the site at −376 relative to the initiator methionine codon (Fig. 2). For describing the 5′-flanking region of the round-spotted pufferfish *JAK1* gene, we used a numbering scheme that the transcription start site is designated +1 (Fig. 3).

We also determined the 2.9 kb sequence of the 5′-upstream region of the *JAK1* gene relative to the transcription start site (Fig. 3). This region was examined for potential DNA elements that may contribute to the round-spotted pufferfish *JAK1* gene regulation and transcription initiation. Computer analysis of the sequence revealed numerous potential binding sites for transcription factors. Two TATA boxes (TATAAA), which is generally located at a position about −30 relative to the RNA start site [13], are observed at −1240 and −2424. Thus, the round-spotted pufferfish *JAK1* gene promoter belongs to the subclass of TATA-less RNA polymerase II promoters which are found in some protein tyrosine kinase genes [14–16]. This gene also has three CCAAT motifs at −2017, −1938, and −1382, which are remote from the standard positions between −60 and −80 relative to the RNA start site [17]. Three putative binding sites for HNF-5 [18] are present at −2910, −705, and −628. Another potential binding site for ubiquitously expressed transcription factors such as NF-IL6 [19] is also observed at −649, −592, and −496. Moreover, the proto-oncogene product E2A [20] binding sites are found at −1336 and −959. There are numerous GC-rich sequences that constitute the potential sites for Sp1 [21] and GCF [22], which are present at −2799, −2688, −2653, −2629, −2606, −1860, and −1352. The GC factor (GCF) has been reported to negatively regulate the expression of epidermal growth factor receptor [22]. The flanking region also contain DNA motifs for CRE [23] at −2572, AP2 [24] at −1821, GATA [25] at −1762, GAGA [26,27] at −1690, and p53 [28] at −933 and −446. While sequencing this 2.9 kb DNA fragment, a $(CA)_{16}$ repeat was found and located at −55 and −24 relative to the transcription initiation site

To verify whether the 5′-flanking region of the round-spotted pufferfish *JAK1* gene exhibits functional promoter activity, the genomic DNA fragment containing the 5′-upstream region (−2922 to +100) was fused to the CAT reporter gene in pCAT–Basic (Promega) to create pRSP1–CAT. Following transfection into carp CF cells [29], the CAT activity of pRSP1–CAT was about 3 and 0.18 times the promoter activity of pJP1–CAT [5] and pRSV–CAT

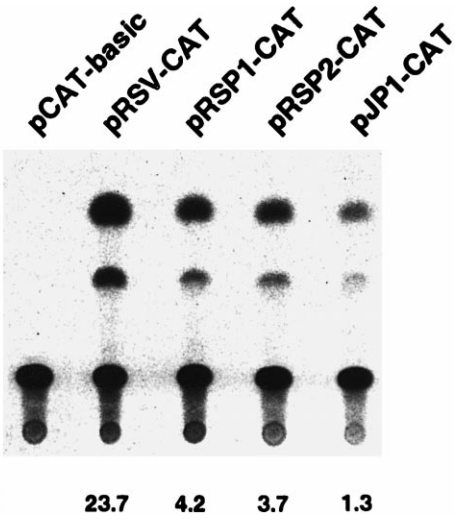

Fig. 4. Analysis of the promoter activity of the 5′-flanking region of the round-spotted pufferfish *JAK1* gene fused to the CAT reporter gene. Each chimeric gene was cotransfected with pSV–β-galactosidase DNA into CF cells and assayed for CAT and β-galactosidase activities as previously described [5]. CAT activity in an individual experiment was corrected for variation in transfection efficiency by normalizing the value to the β-galactosidase activity in the same extract. The data represented the mean of triplicate transfection experiments for each plasmid. The acetylated products of the CAT assay were separated by thin layer chromatography developed with chloroform–methanol (95:5, v/v), visualized by autoradiography and quantitated by using the PhosphoImager (Bio-Imaging Analyzer BAS 2000, Fuji, Japan).

[30], respectively (Fig. 4). Therefore, the promoter region of the round-spotted pufferfish *JAK1* gene displays stronger activity than that of the common carp *JAK1* promoter even in a carp cell line. As shown in Fig. 3, there are three potential sites (at −2799, −2688, and −2629) for GCF, which has been reported to be a negatively regulator in the expression of epidermal growth factor receptor [22]. In order to investigate the effect of these sites on the promoter activity of the round-spotted pufferfish *JAK1* gene, a deletion mutant, pRSP2–CAT (nucleotides −2484 to +100), was constructed from pRSP1–CAT. However, both pRSP1–CAT and pRSP2–CAT had almost the same activity as shown in Fig. 4. This result needs further investigation to analyze the endogenous expression level of GCF in this carp cell line.

The pufferfish *Fugu rubripes* (*Fugu*) has been used as a model for vertebrate genome analysis [7,9]. Its genome is estimated to be approximately 404 Mb, 7.5 times smaller than that of human. Although the relative gene order in the AD3 locus is the same in *Fugu* and human genomes [31], whereas the relative gene order in the *Surfeit* locus, which contains *Surfeit*1 to *Surfeit*6, is different between Fugu and mammals [32]. However, *Fugu* homologues of six *Surfeit* genes are all highly conserved at the amino acid level and their gene structures are mostly identical to the mammalian genes. In addition, the genome sizes in the AD3 locus are 12.4 kb in *Fugu*, compared to > 600 kb in human [31]. Therefore, it has been suggested that *Fugu* genome could be used to identify human disease genes using comparative mapping or positional cloning.

In this study, we have cloned and characterized the *JAK1* gene from the round-spotted pufferfish *Tetraodon fluviatilis*, whose genome size is 4 and 8 times smaller than that of common carp and human [7,8], respectively. The exon–intron organization of both fish *JAK1* genes were compared (data not shown). All of the splice sites of both genes are identical except intron 6, which is only present in the round-spotted pufferfish *JAK1* gene. Comparison of the intron sizes of both *JAK1* genes are shown in Table 2. There are only three introns larger than 1 kb for the round-spotted pufferfish *JAK1* whereas there are 11 larger introns (> 1 kb) for the common carp *JAK1*. Based on the comparison of genomic organization and intron sizes in the *JAK1* gene, our results indicate that the round-spotted pufferfish is also a good model organism for comparative vertebrate genomic structure analysis.

Table 2
Comparison of intron size (bp) of *JAK1* genes

Common carp	Round-spotted pufferfish
> 3000	1646
1800	637
1000	1403
400	228
148	106
0	200
350	106
700	117
1200	201
1500	129
2300	1638
900	137
187	128
2200	78
1500	120
820	710
111	75
457	450
110	141
100	68
1800	386
170	93
1500	195
1400	122

The first intron of fish *JAK1* genes interrupts the corresponding 5′-untranslated region of the *JAK1* mRNA sequence and the sizes of this intron are quite large, 1646 bp for the round-spotted pufferfish *JAK1* and more than 3 kb for the common carp *JAK1*. Such a phenomenon that a large 1st intron followed by a small untranslated region has been found among many protein tyrosine kinase genes, such as human *lck* gene [33], mouse *hck* gene [34], and human α-platelet-derived growth factor receptor gene [16]. At present, due to the unclonable nature of most part of the 1st intron, the accurate size of the 1st intron of the common carp *JAK1* gene has not been determined [5]. However, through sequencing and computer analysis, it is interesting to find a putative STAT-binding sequence (5′-TTCCGTGAA-3′) in the 1st intron of the round-spotted pufferfish *JAK1*. The significance of the presence of putative STAT-binding motif in the large 1st intron awaits further investigation.

Although the compact genome of *Fugu* has been proposed to be suited for detecting conserved regulatory elements present in the noncoding region [35], our studies indicate that the promoter region of the round-spotted pufferfish *JAK1* gene is different from that of the carp *JAK1* gene. In carp, there are several potential transcription factor binding sites, such as AP1, E2A, GHF-5, HNF-5, and NF-IL6 in the *JAK1* 5′-flanking region [5]. On the contrary, the putative promoter region of round-spotted pufferfish *JAK1* gene not only has E2A, HNF-5, and NF-IL6 sites, but also has other potential binding sites for transcription factors, such as CRE, GCF, GAGA, GATA, p53, and Sp1 (Fig. 3). Moreover, the CAT activity of the 2.9 kb DNA fragment of the 5′ flanking of the round-spotted pufferfish *JAK1* gene was 3-fold higher than that of the 2.5 kb DNA fragment of the 5′ flanking of the carp *JAK1* gene as transfection into a carp cell line (Fig. 4). The relevance of CAT activity to the presence of other potential regulatory elements in the 5′ flanking region of the round-spotted pufferfish *JAK1* gene needs further investigation.

This research was supported by grants from the National Science Council (NSC-85-2311-B-001-046) and from the Department of Health (DOH85-HR-507), Taiwan.

References

[1] A.F. Wilks, A.G. Harpur, R.R. Kurban, S.J. Ralph, G. Zurcher, A. Ziemiecki, Mol. Cell. Biol. 11 (1991) 2057–2065.
[2] J.N. Ihle, I.M. Kerr, Trends Genet. 11 (1995) 69–74.
[3] C. Schindler, J.E. Darnell Jr., Annu. Rev. Biochem. 64 (1995) 621–651.
[4] R. Binari, N. Perrimon, Genes Dev. 8 (1994) 300–312.
[5] M.S. Chang, G.D. Chang, J.H. Leu, F.L. Huang, C.K. Chou, C.J. Huang, T.B. Lo, DNA Cell Biol. 15 (1996) 827–844.
[6] X.S. Hou, M.B. Melnick, N. Perrimon, Cell 84 (1996) 411–419.
[7] S. Brenner, G. Elgar, R. Sandford, A. Macrae, B. Venkatesh, S. Aparicio, Nature (Landon) 366 (1993) 567–571.
[8] R. Hinegardner, D.E. Rosen, Am. Nat. 106 (1972) 621–644.
[9] G. Elgar, R. Sandford, S. Aparicio, A. Macrae, B. Venkatesh, S. Brenner, Trends Genet. 12 (1996) 145–150.
[10] J. Sambrook, E.F. Fritsch, T. Maniatis, in: N. Irwin, N. Ford, C. Nolan (Eds.), Molecular Cloning, A Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989.
[11] R. Breathnach, C. Benoist, K. O'hare, F. Gannon, P. Chambon, Proc. Natl. Acad. Sci. U.S.A. 75 (1978) 4853–4857.
[12] X. Yang, D. Chung, C.L. Cepko, J. Neurosci. 13 (1993) 3006–3017.
[13] R. Breathnach, P. Chambon, Annu. Rev. Biochem. 50 (1981) 349–383.
[14] M. Patel, S.J. Leevers, P.M. Brickell, Oncogene 5 (1990) 201–206.
[15] U. Lichtenberg, N. Quintrell, J.M. Bishop, Oncogene 7 (1992) 849–858.
[16] J. Kawagishi, T. Kumabe, T. Yoshimoto, T. Yamamoto, Genomics 30 (1995) 224–232.
[17] L.A. Chodosh, A.S. Baldwin, R.W. Carthew, P.A. Sharp, Cell 53 (1988) 11–24.
[18] T. Grange, J. Roux, G. Rigaud, R. Pictet, Nucleic Acids Res. 19 (1991) 131–139.
[19] S. Akira, H. Isshiki, T. Sugita, O. Tanabe, S. Kinoshita, Y. Nishio, T. Nakajima, T. Hirano, T. Kishimoto, EMBO J. 9 (1990) 1897–1906.
[20] C. Murre, P.S. McCaw, D. Baltimore, Cell 56 (1989) 777–783.
[21] M.R. Briggs, T. Kadonaga, S.P. Bell, R. Tjian, Science 234 (1986) 47–52.
[22] R. Kageyama, I. Pastan, Cell 59 (1989) 815–825.
[23] G.A. Gonzalez, M.R. Montminy, Cell 59 (1989) 675–680.
[24] S. Faisst, S. Meyer, Nucleic Acids Res. 20 (1992) 3–26.
[25] T. Evans, M. Reitman, G. Felsenfeld, Proc. Natl. Acad. Sci. U.S.A. 85 (1988) 5976–5980.
[26] M.D. Biggin, R. Tjian, Cell 53 (1988) 699–711.
[27] W.C. Soeller, S.J. Poole, T.B. Kornberg, Genes Dev. 2 (1988) 68–81.
[28] S.E. Kern, K.W. Kinzler, A. Bruskin, D. Jarosz, P. Friedman, C. Prives, B. Vogelstein, Science (Washington, D.C.) 252 (1991) 1708–1711.
[29] S.N. Chen, G.H. Kou, in: E. Kurstak, Y. Kuroda (Eds.), Invertebrate and Fish Tissue Culture, Japan Societies Press, Springer, Berlin, 1986, pp. 218–227.
[30] C. Gorman, C. Padmanabhan, B.H. Howard, Science (Washington, D.C.) 221 (1983) 551–553.
[31] M.K. Trower, S.M. Orton, I.J. Purvis, P. Sanseau, L. Riley, C. Cristodoulou, D. Burt, C.G. See, G. Elgar, R. Sherrington, E.I. Rogaev, P.S. George-Hyslop, S. Brenner, C.W. Dykes, Proc. Natl. Acad. Sci. U.S.A. 93 (1996) 1366–1369.
[32] J. Gilley, N. Armes, M. Fried, Nature (London) 385 (1997) 305–306.
[33] E. Rouer, T.V. Huynh, S. Lavareda de Souza, M.-C. Lang, S. Fischer, R. Bernarous, Gene 84 (1989) 105–113.
[34] S.F. Ziegler, C.M. Pleiman, R.M. Perlmutter, Oncogene 6 (1991) 283–288.
[35] S. Aparicio, A. Morrison, A. Gould, J. Gilthorpe, C. Chaudhuri, P. Rigby, R. Krumlauf, S. Brenner, Proc. Natl. Acad. Sci. U.S.A. 92 (1995) 1684–1688.