Molecular characterization and subcellular localization of *Carassius auratus* interferon regulatory factor-1

Yan Shi¹, Yi-Bing Zhang¹, Zhe Zhao, Jun Jiang, Qi-Ya Zhang, Jian-Fang Gui

State Key Laboratory of Freshwater Ecology and Biotechnology, Institute of Hydrobiology, Chinese Academy of Sciences, Graduate School of the Chinese Academy of Sciences, Wuhan 430072, China

Received 7 March 2007; received in revised form 8 May 2007; accepted 8 May 2007

Available online 14 June 2007

**KEYWORDS**
Interferon regulator factor-1 (IRF-1); *Carassius auratus*; Grass carp hemorrhagic virus; Inductive expression; Nuclear localization signal (NLS); Subcellular localization

**Abstract**
Interferon (IFN)-regulated transcription factor-1 (IRF-1) has been studied in mammals and fish, but little is known about the relationship between its gene structure and nuclear localization of IRF-1 protein. In this study, a cDNA encoding *Carassius auratus* IRF-1 (CaIRF-1) was isolated from an interferon-producing cell line, *C. auratus* blastulae embryonic (CAB) cells, exposed to UV-inactivated grass carp hemorrhagic virus (GCHV). The CaIRF-1 genomic locus exhibits exon–intron arrangements similar to those of other vertebrate IRF-1 loci, with nine exons and eight introns, although together with pufferfish IRF-1, CaIRF-1 distinguishes itself from other vertebrate IRF-1 genes by a relatively compact genomic size. Similar to the known IRF-1 genes, CaIRF-1 is ubiquitously expressed, and is upregulated in vitro and in vivo in response to virus, Poly I:C, or CAB IFN-containing supernatant (ICS). Subcellular localization analysis confirms the nuclear distribution of CaIRF-1 protein, and reveals two nuclear localization signals (NLS), any one of which is sufficient for nuclear translocation of CaIRF-1. One NLS locates to amino acids 117–146, and appears to be the structural and functional equivalent of the NLS in mammalian IRF-1. The second NLS (amino acids 73–115) is found within the DNA-binding domain (DBD) of CaIRF-1, and contains two regions rich in basic amino acids ("95KDKSINK101" and "75KTWKANFR82"). In comparison with mammalian IRF-1, in which the corresponding amino acid stretch does not seem to drive nuclear translocation, five conserved basic amino acids (K75, K78, R82, K95, and K101) and one non-conserved basic amino acid (K97) are present in this NLS from CaIRF-1. This observation suggests that K97 of CaIRF-1 might be essential for the function of its second NLS, wherein the six basic amino

---

The nucleotide sequence data reported in this paper have been submitted to the GenBank under accession number **EF174419** for CaIRF-1 cDNA, and **EF120633** for CaIRF-1 genomic DNA.

Corresponding author. Tel.: +86 27 68780707; fax: +86 27 68780123.

E-mail address: jfgui@ihb.ac.cn (J.-F. Gui).

¹The two authors contributed equally to this work.
1. Introduction

The interferon-regulatory factor (IRF) family comprises transcription factors that regulate the expression of interferon (IFN) and IFN-stimulated genes (ISGs) by binding to characteristic elements in their promoters [1,2]. Among IRFs, at least four members, IRF-1, IRF-3, IRF-5, and IRF-7, have been implicated as direct transducers of virus-mediated IFN signaling [3]. IRF-1 was the first IRF family member known to activate the IFN-β gene, and found to be constitutively expressed in most cell types [4]. IRF-1 is dramatically upregulated upon virus infection, IFN stimulation, or treatment with a number of cytokines such as interleukin-1, tumor necrosis factor, platelet-derived growth factor, and colony stimulating factor [2]. In mammals, IRF-1 is indispensable for the antiviral action of IFNs against some viruses, including EMCV [5,6], and is one of the key host factors to regulate intracellular HCV replication through the modulation of ISG-mediated antiviral responses [7]. In some instances though, the role of IRF-1 in IFN-dependent responses is less clear, as exemplified by the normal levels of type I IFN transcripts in IRF-1-deficient embryonic fibroblasts (MEFs) infected with Newcastle disease virus (NDV) [8]. In addition to the regulation of IFN, IRF-1 expression is linked to the control of cell growth and apoptosis in a p53-dependent or p53-independent manner [9,10]. Recent evidence, for instance, has shown that overexpression of IRF-1 is associated with growth suppression of human breast cancer cells [11,12]. IRF-1 selectively modulates different sets of genes, depending on the cell type and/or the nature of cellular stimuli, to evoke appropriate responses, such as T-helper-1-mediated immune response, natural killer activity, antiproliferation, apoptosis, antiviral activity, and oncogenesis [2].

In mammals, nine IRF family members have been identified [3]. The N-terminal 115 amino acids are strongly conserved in all members of the family, and comprise the DNA-binding domain (DBD). This DBD displays a unique helix-turn-helix structure, with the DNA recognition helix defined by a cluster of five highly conserved tryptophan residues [13]. Some IRF family members, including IRF-3, IRF-4, IRF-5, IRF-8, and IRF-9, also share a sequence homologous to the transactivation domain (TAD) of Smad morphogens at their C-termini, which is termed the IRF association domain (IAD) [14]. While the IRF-1 C-terminal region also possesses a TA, its sequence is much more divergent when compared with those of other IRF family members [15–17].

Constitutive nuclear localization of mammalian IRF-1 has been confirmed [4,18], and its nuclear localization signal (NLS) has been localized to a region C-terminal of the DBD, spanning amino acids 117–141 [16]. Generally, an NLS consists of a stretch of lysines (K) and arginines (R), but this is not strictly conserved among IRF family members [19]. A bipartite NLS has also been identified within the N-terminal DBD of IRF-9 and IRF-4 [20,21]. In addition, a nuclear export sequence (NES) determining cytoplasmic localization has been defined in two IRF proteins, IRF-3 and IRF-5 [22,23].

Japanese flounder (Paralichthys olivaceus) IRF-1 was the first fish homologue of IRF-1 to be identified [24]. Since then, different IRF genes have been cloned in pufferfish (Fugu rubripes) [25], rainbow trout (Onchorhynchus mykiss) [26], crucian carp (Carassius auratus L.) [27], zebrafish (Danio rerio) [28], turbot (Scophthalmus maximus), and sea bream (Sparus aurata) [29]. A recent study has demonstrated the ability of flounder IRF-1 to induce an antiviral state in cells, by mediating the production of cytokine-like substances [30]. However, little is currently known about the relationship between structure and nuclear localization in any fish homologue of IRF-1. To help understand the antiviral functions of fish IRF-1, we characterized C. auratus IRF-1 (CaIRF-1) and analyzed its genomic structure. Various CaIRF-1 mutants fused to GFP were constructed, and transient transfection assays were carried out. As a result of these studies, two NLSs were identified in CaIRF-1, which is in contrast with the single NLS found in mammalian IRF-1 proteins. Furthermore, site-directed mutagenesis experiments defined six basic amino acids localized within residues 75–113 of the DBD, which contribute to the nuclear translocation of CaIRF-1. These novel findings provide insight into the structure–function relationships that determine nuclear localization of fish IRF-1, and thus into its antiviral roles.

2. Materials and methods

2.1. Cells, virus, and inductions

Crucian carp (C. auratus L.) blastulae embryonic cells (CAB) and grass carp (Ctenopharyngodon idellus) ovary cells (CO) were maintained in medium 199 supplemented with 10% fetal calf serum (FCS), 100 U/ml penicillin, and 100 μg/ml streptomycin. Flounder (P. olivaceus) embryonic cells (FEC) were cultured in Dulbecco’s modified Eagle’s medium with the same concentration of FCS and antibiotics mentioned above. Two dsRNA viruses, grass carp hemorrhagic virus (GCHV) and flounder birnavirus (FBV), were propagated in CAB and FEC cells, respectively. UV inactivation and titer determination were executed as described previously [31]. Briefly, the harvested cell culture fluid containing GCHV was centrifuged at 4 × 10⁶ g for 20 min to remove the cell debris, and the supernatant was then ultracentrifuged at 1 × 10¹³ g for 1.5 h. The virus pellet was obtained and resuspended in optimal medium 199 without FCS, and centrifuged again at 1.5 × 10¹³ g at 4°C for 20 min. A volume of 3.5–4 ml (1 × 10⁶ TCID₅₀/ml) of purified GCHV inside a 40-mm Petri dish was placed under constant and slow shaking, and exposed to UV irradiation for 5 min with a 30 W General
Electric germicidal lamp placed at a distance of 15 cm. The effectiveness of virus inactivation was determined by the 50% tissue culture infectious dose (TCID$_{50}$) assay according to a previous report [27].

Induction of CAB cells was performed following the procedures described previously [31]. Briefly, after 2 days of culture, CAB cells were treated for 1 h with 0.5 ml UV-inactivated GCHV ($1 \times 10^7$ TCID$_{50}$/ml exposed to UV irradiation), active FBV ($1 \times 10^7$ TCID$_{50}$/ml), Poly I:C (100 $\mu$g/ml, Sigma), or CAB ICS (2000 U/ml). Treated cells were then supplemented with appropriate FCS-free 199 medium to a total volume of 5 ml and further cultured at 28 $^\circ$C. The cells were collected for total RNA extraction at 2, 6, 12, 24, 48, 72, 96, and 120 h after treatment. Control cells were treated with FCS-free 199 medium alone followed by collection at 120 h. Crucian carp weighing between 30 and 45 g were divided into three groups, each with three fishes. One group was injected abdominally with active GCHV ($1 \times 10^7$ TCID$_{50}$/ml, 500 ml per fish), the second group was injected with Poly I:C (100 mg/ml, Sigma), or CAB ICS (2000 U/ml), and the third group received PBS as a control (500 ml per fish). At 24 h post injection, tissues were isolated for RNA extraction.

2.2. Transfection and subcellular localization

Transfections were carried out using Lipofectamine 2000 reagent (Invitrogen). Before transfection, CAB cells or CO cells were grown overnight to 90% confluence in six-well plates, and 4 $\mu$g of DNA and 10 $\mu$l of liposome reagent were pre-mixed in FCS-free 199 medium followed by 20 min of incubation. After the medium was removed, cells in each well were incubated with 500 $\mu$l of the DNA:liposome mixture. Six hours later, the mixture was substituted by fresh 199 medium and the transfected cells were directly observed for subcellular localization of CaIRF-1. For staining with propidium iodide (PI), the cells transfected for 24 h were rinsed with PBS three times, fixed with 4% (v/v) paraformaldehyde for 20 min at room temperature, incubated with 0.2% Triton X-100 for 30 min, and finally stained with 1 $\mu$g/ml PI for 1 h away from light. All samples were examined under a Leica DM IRB fluorescence microscope.

2.3. RNA extraction, reverse transcription, and genomic DNA isolation

Total RNA from CAB cells and crucian carp tissues were extracted by TRIZOL Reagent (Invitrogen) and SV Total RNA Isolation System (Promega), respectively. First-strand cDNA was synthesized using random primers and M-MLV reverse transcriptase (Promega) according to the procedure described previously [32]. Genomic DNA was isolated by standard phenol/chloroform extraction.

2.4. RACE-PCR and real-time PCR

Primers ZeIRF1-F and ZeIRF1-R, derived from the conserved DNA-binding domain of the zebrafish IRF-1 gene (AY398364), were used to amplify the corresponding CaIRF-1 fragment from a SMART cDNA library made with mRNA from CAB cells treated with UV-inactivated GCHV [33]. Subsequently, the full-length cDNA of CaIRF-1 was obtained by RACE-PCR. The 5’ sequence was amplified with primers SMART-F and ZeIRF1-R, and the 3’ end was amplified with primers ZeIRF1-F and SMART-R. Standard PCR was used to clone genomic DNA segments of CaIRF-1 with two pairs of primers, ZeIRF1-F and ZeIRF1-R, and IRF1D-F and IRF1D-R, using CAB cell genomic DNA as template. Primers ZeIRF1-F and IRF1D-R were used to amplify the full-length CaIRF-1 locus for confirming the continuity of the genomic DNA sequence. All PCR primers used throughout this study are shown in Table 1.

Real-time PCR was used to analyze the expression of CaIRF-1 using primers IRF1RT-F and IRF1RT-R on a DNA Engine ARTICLE IN PRESS

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Sequences (5’–3’)</th>
<th>Usage</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZeIRF1-F</td>
<td>ATGCCCGTGTCCAGAATGC</td>
<td>Gene cloning</td>
</tr>
<tr>
<td>ZeIRF1-R</td>
<td>TCCTTCACTCCTCGATATCCG</td>
<td>Genomic DNA cloning</td>
</tr>
<tr>
<td>IRF1D-F</td>
<td>TTGGTACGGAGGACACATGTC</td>
<td>Plasmid construction</td>
</tr>
<tr>
<td>IRF1D-R</td>
<td>GCATCGCTAGCTACTAACCATGCTG</td>
<td></td>
</tr>
<tr>
<td>IRF1G-F1</td>
<td>ATGGTACCCGAGGACACATGTC</td>
<td></td>
</tr>
<tr>
<td>IRF1G-R1</td>
<td>TCAAGCTTCATGACGAAAGAACAA</td>
<td></td>
</tr>
<tr>
<td>IRF1G-F2</td>
<td>CACGGTAGCTCTTGGCTGAGAC</td>
<td></td>
</tr>
<tr>
<td>IRF1G-R2</td>
<td>TCTGGTGAGCCGGAGCCGC</td>
<td></td>
</tr>
<tr>
<td>IRF1G-R3</td>
<td>TCTGGTGAGCCGGAGCCGC</td>
<td></td>
</tr>
<tr>
<td>IRF1G-R4</td>
<td>TCTGGTGAGCCGGAGCCGC</td>
<td></td>
</tr>
<tr>
<td>IRF1RT-F</td>
<td>CGGCTGAGAGGACGAGGACAA</td>
<td>Real-time PCR</td>
</tr>
<tr>
<td>IRF1RT-R</td>
<td>GCTGAGAGGACGAGGACAA</td>
<td></td>
</tr>
<tr>
<td>CaMx1-F</td>
<td>ACAGAAAGAAGGAGGAGGAGGTA</td>
<td></td>
</tr>
<tr>
<td>CaMx1-R</td>
<td>CAGGAGTTCTCAGAAGC</td>
<td></td>
</tr>
<tr>
<td>SMART-F</td>
<td>AACCGAGAGTAGGGGG</td>
<td>RACE-PCR</td>
</tr>
<tr>
<td>SMART-R</td>
<td>CAGGAGATC16</td>
<td></td>
</tr>
<tr>
<td>$\beta$-Actin F</td>
<td>CACTCTGCCCATCTACAG</td>
<td>Real-time PCR</td>
</tr>
<tr>
<td>$\beta$-Actin R</td>
<td>CACTCTGCCCATCTACAG</td>
<td></td>
</tr>
</tbody>
</table>

Table 1 Primers used for all the experiments
Chrom 4 Real-time System (MJ Research). Reactions were performed in a 20 μl volume containing SYBR Green I Dye. Cycling parameters were 94°C for 4 min, followed by 40 cycles of 94°C for 20 s, 52°C for 20 s, and 72°C for 20 s. β-actin cDNA was amplified in parallel in order to normalize the template concentrations. All samples were analyzed in triplicate, and the results were described as expression units relative to the β-actin gene in each sample using the 2^(-ΔΔC(T)) method [34].

2.5. Plasmid construction and mutagenesis

Seven fusion plasmids, IRF-1-GFP (amino acids 1–289), NLS-GFP (116–146), ΔC-GFP (1–146), AN-GFP (116–289), N-GFP (1–115), C-GFP (147–289), and M-GFP (1–72), were generated by fusing different PCR fragments of CalIF-1 with in-frame restriction sites into pEGFP-N3 vector (Clontech). An internal deletion mutant, ΔNLS-GFP (111–146), was constructed by two rounds of PCR, performed with two sets of specially designed primers using codon degeneracy. Briefly, both the 3’ end of the PCR fragment spanning amino acids 1–115 and the 5’ end of the PCR fragment spanning amino acids 147–289 contained Sal I restriction sites that were integrated using codon degeneracy. The two PCR products were digested with Sal I, joined with ligase, and finally subcloned into pEGFP-N3 vector with in-frame Nhe I and Kpn I restriction sites. Two rounds of site-directed mutagenesis were performed with the QuikChange II Site-Directed mutagenesis kit (Stratagene); K95, K97, and K101 of N-GFP (1–115) were changed to alanine to create NA1-GFP, and further mutation of K75, K78, and R82 resulted in plasmid NA2-GFP (where K95, K97, K101, K75, K78, and R82 were changed to A). All constructs were verified by sequencing.

2.6. Sequence alignments and phylogenetic analysis

Multiple alignments were performed with CLUSTALW 1.83 [35], and used to derive a phylogenetic tree by neighbor-joining methods. The homologous sequences were searched using the National Center for Biotechnology Information blast server [36]. NLS prediction was performed using PSORTII web server (http://psort.nibb.ac.jp/form2.html).

3. Results

3.1. Molecular characterization of CalIF-1

The full-length cDNA of CalIF-1 was generated by the 3’ and 5’ RACE methods, and comprises 1450 bp. The longest open reading frame (ORF) encodes a 289-amino acid protein, with a calculated molecular weight of 32.67 kDa and a pI of 5.88. In the 3’ untranslated region, an AU sequence (ATTTA) is found, a motif possibly involved in mRNA instability [37]. A computer search showed CalIF-1 to be highly similar to IRF-1 genes deposited in public databases, with zebrafish IRF-1 as its closest relative.

Multiple alignments revealed that, like mammalian IRF-1 proteins, the putative CalIF-1 protein has a DBD spanning the N-terminal 113 amino acids, including five characteristic tryptophan residues. The N-terminal DBDs of IRF-1 proteins represent regions of higher conservation than their C-terminal sequences (Fig. 1). However, within the C-terminus of CalIF-1, there is a relatively high level of sequence conservation in the putative core TAD, which has been identified recently by transactivation analysis of human IRF-1 [17]. CalIF-1 contains a basic K- and R-rich sequence (117KKKDKRPKGRDSRRVKK141) located in the region equivalent to the NLS of mammalian IRF-1 proteins. NLS prediction by PSORTII suggests that this stretch contains a bipartite NLS. CalIF-1 is overall 36.6–76.2% identical to homologous proteins from mammals, birds, amphibians, and fish, with 72.6–95.6% identity at the level of the N-terminal DBD and only 18.4–66.8% identity in the C-terminus (Table 2).

The high similarity of CalIF-1 to known members of the IRF-1 subfamily was further supported by a phylogenetic tree, which was built with 18 IRF-1 proteins and 8 IRF-2 proteins from mammals, birds, amphibians, and fish. As shown in Fig. 2, three distinct classes are retrieved, where one cluster consists of only IRF-2 proteins, and IRF-1 proteins are divided into two distinct clusters of fish and non-fish factors. This finding implies independent divergence of IRF-1 and IRF-2 after separation of fish from other vertebrates.

3.2. Genomic organization of CalIF-1

To further explore the evolutionary conservation of IRF-1 genes, a 1570 bp genomic DNA sequence covering the whole ORF encoded in CalIF-1 was amplified. Based on the CalIF-1 cDNA sequence, the exon/intron organization of the CalIF-1 gene was investigated. Computer searches of public genomic databases and subsequent comparison between cDNA and genomic DNA sequences also revealed the sizes and organizations of IRF-1 genes from human, mouse, rat, cattle, chick, and pufferfish. As shown in Fig. 3, all IRF-1 genomic loci have nine exons and eight introns, and the locations of splice donor/acceptor sites in all introns follow the consensus ‘GT/AG’ rule [38] (Table 3). Interestingly, two fish IRF-1 genes, crucian carp CalIF-1 (1.57 kb) and pufferfish TriRF-1 (1.73 kb), show a compact genomic size that is much shorter than any IRF-1 from non-fish vertebrate species (4.14–5.53 kb). However, the size and coding potential of exons that represent the common regions of IRF-1 proteins have been conserved in the fish homologues. For example, similar to human and mouse IRF-1 genes [16], exons 1, 2, and 3 of the CalIF-1 gene encode the N-terminal DBD (corresponding to amino acids 1–120).

3.3. Inductive expression of CalIF-1 by virus, CAB ICS or Poly I:C

In mammals, IRF-1 is induced by IFN, virus, and Poly I:C [2]. In order to determine whether CalIF-1 has a similar expression pattern, CalIF-1 mRNA levels were examined in CAB cells after exposure to CAB IFN-containing supernatant (ICS), two kinds of viruses, GCHV and active FBV, or treatment with Poly I:C. For the induction of CAB cells, UV-inactivated GCHV rather than active GCHV was used due to its higher effectiveness in inducing IFN activity [31].
**Fig. 1** Multiple alignment of CaIRF-1 amino acid sequence with other IRF-1 proteins from fish, human, mouse, rat, and chick. Missing amino acids are denoted by hyphens. The DNA-binding domain (DBD) and transactivation domain (TAD) are indicated by lines above the aligned sequences. The conserved tryptophan (W) residues that comprise a “tryptophan cluster” are highlighted by asterisks. The bipartite NLS of mammalian IRF-1 proteins is boxed. The second NLS in CaIRF-1 is located within amino acids 73–115, and the six basic amino acids essential for its function are indicated by open triangles. The alignment was refined using GeneDoc software. Identical amino acid residues, and those sharing biochemical properties in all sequences are highlighted with black shading, while gray shading was used for regions with more than 60% identity, or where more than 9 of 14 amino acid residues shared the same properties.
Table 2  Pairwise identity of CaIRF-1 with other selected IRF-1 proteins

<table>
<thead>
<tr>
<th>Gene</th>
<th>Species</th>
<th>GenBank accession number</th>
<th>Identity percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Whole DBD C-terminal</td>
</tr>
<tr>
<td>Teleosts</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CaIRF-1</td>
<td>Crucian carp (Carassius auratus L.)</td>
<td>EF174419</td>
<td>76.2 95.6 66.8</td>
</tr>
<tr>
<td>DrIRF-1</td>
<td>Zebrafish (Danio rerio)</td>
<td>AY398364</td>
<td></td>
</tr>
<tr>
<td>OmiIRF-1</td>
<td>Rainbow trout (Oncorhynchus mykiss)</td>
<td>AF332147</td>
<td>54.5 81.4 41.5</td>
</tr>
<tr>
<td>SmIRF-1</td>
<td>Turbot (Scophthalmus maximus)</td>
<td>AY962253</td>
<td>51.2 81.4 36.4</td>
</tr>
<tr>
<td>SmIRF-1</td>
<td>Turbot (Scophthalmus maximus)</td>
<td>AY962252</td>
<td>51.2 81.4 36.4</td>
</tr>
<tr>
<td>SAILR-1</td>
<td>Gilthead seabream (Sparus aurata)</td>
<td>AY962254</td>
<td>50.7 77 37.9</td>
</tr>
<tr>
<td>SAILR-1</td>
<td>Gilthead seabream (Sparus aurata)</td>
<td>AY962256</td>
<td>53.7 77 33.6</td>
</tr>
<tr>
<td>ScIRF-1</td>
<td>Chinese perch (Siniperca chuatsi)</td>
<td>AY647431</td>
<td>50.3 79.6 35.9</td>
</tr>
<tr>
<td>PoIRF-1</td>
<td>Flounder (Paralichthys olivaceus)</td>
<td>AB005883</td>
<td>50.9 80.5 36.4</td>
</tr>
<tr>
<td>TrIRF-1</td>
<td>Pufferfish (Takifugu rubripes)</td>
<td>AF242474</td>
<td>50.3 73.5 39</td>
</tr>
<tr>
<td>Avians</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GgIRF-1</td>
<td>Chicken (Gallus gallus)</td>
<td>AJ621385</td>
<td>40.1 76.1 22.6</td>
</tr>
<tr>
<td>CcIRF-1</td>
<td>Common quail (Coturnix coturnix)</td>
<td>AJ271052</td>
<td>39.5 76.1 21.8</td>
</tr>
<tr>
<td>Mammals</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OaIRF-1</td>
<td>Sheep (Ovis aries)</td>
<td>AF331970</td>
<td>39.8 74.3 23.1</td>
</tr>
<tr>
<td>BtIRF-1</td>
<td>Cattle (Bos taurus)</td>
<td>AY204899</td>
<td>39.5 74.3 22.6</td>
</tr>
<tr>
<td>HsIRF-1</td>
<td>Human (Homo sapiens)</td>
<td>BT019756</td>
<td>38.9 74.3 21.8</td>
</tr>
<tr>
<td>MmIRF-1</td>
<td>Mouse (Mus musculus)</td>
<td>NM_008390</td>
<td>36.6 74.3 18.4</td>
</tr>
<tr>
<td>RnIRF-1</td>
<td>Rat (Rattus norvegicus)</td>
<td>BC076382</td>
<td>37.2 74.3 19.2</td>
</tr>
<tr>
<td>Amphibians</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>XtIRF-1</td>
<td>Silurana tropicalis (Xenopus tropicalis)</td>
<td>NM_001006694</td>
<td>40.5 72.6 24.9</td>
</tr>
</tbody>
</table>

Fig. 2  Phylogenetic relationships within the IRF-1 protein family. A neighbor-joining phylogenetic tree [46] was constructed based on analysis of protein sequences of the IRF-1 subgroup and the IRF-2 subgroup from fish, birds, and mammals. The bootstrap confidence values shown at the nodes of the tree are based on 1000 bootstrap replications.
Real-time PCR analyses showed that all stimuli induced a significant and different increase in CaIRF-1 mRNA level, whose expression patterns resembled those of CaMx-1, a hallmark gene of the IFN antiviral response (Fig. 4A). Upon stimulation, the mRNA levels of both genes first increased gradually, peaked, and thereafter decreased back to the basal level. Among the four stimuli, CAB ICS and UV-inactivated GCHV showed a greater ability to induce CaMx and CaIRF-1 than FBV and Poly I:C. The two most effective inducers led to high levels of expression of CaIRF-1, with maximal induction at 96 h for CAB ICS (about 130-fold, relative to the basal expression in normal CAB cells), and peak expression at 24 h for UV-inactivated GCHV (about 140-fold induction). Although FBV was unable to produce a cytopathic effect in CAB cells, it nevertheless displayed the ability to activate CaIRF-1 transcription, with maximal expression at 48 h post infection (about 16-fold induction). This peak induction of CaIRF-1 by FBV was greater than the peak induction caused by Poly I:C treatment, which also occurred at 48 h post stimulation (by about 7-fold) (Fig. 4A).

Fig. 4 also shows constitutive expression of CaIRF-1 in normal CAB cells (Fig. 4A) and in various organs of healthy crucian carp (Fig. 4B), including gill, spleen, liver, head kidney, posterior kidney, intestine, brain, and heart. The most predominant expression was seen in gill, spleen, and posterior kidney. After injection with GCHV or Poly I:C, upregulation of CaIRF-1 was observed in gill and head kidney (Fig. 4C).

3.4. Nuclear localization of CaIRF-1

As a transcription factor, human IRF-1 protein is constitutively located in the nucleus [16]. To probe the intracellular localization of CaIRF-1 protein, an expression plasmid was constructed, IRF-1–GFP, which drives the expression of a fusion green fluorescent protein consisting of full-length CaIRF-1 and GFP (Fig. 5). As shown in Fig. 6A, wild-type GFP protein is found both in the nucleus and in the cytoplasm, whereas the fusion protein IRF-1–GFP only accumulates in the nuclear compartment.

A time-course experiment further allowed the characterization of the dynamics of nuclear retention of CaIRF-1 protein in CAB cells and CO cells (Fig. 6B). Soon after transfection, the IRF-1–GFP fusion protein started to be expressed and was ubiquitously distributed throughout the nucleus. With the increase in the amount of synthesized IRF-1–GFP fusion protein over time, the green fluorescent signal became compact, and accumulated gradually as green spots that further converged and compacted inside the
nucleus. In comparison with CAB cells, CO cells showed higher transfection efficiency (data not shown), which probably explains the faster and stronger green fluorescent signal noted in these cells, where it was easier to observe the nuclear accumulation of CaIRF-1 (Fig. 6B).

3.5. Identification of functional NLSs in CaIRF-1

Members of the IRF family such as IRF-1, IRF-2, IRF-3, IRF-4, IRF-5, IRF-8, and IRF-9 are known to possess functional NLSs [16,19–21,23]. To confirm whether the sequence...
**Ca IRF-1 functions as an NLS, a series of Ca IRF-1 mutants fused to GFP were prepared (Fig. 5) and tested in both CAB cells and CO cells (Fig. 7).** As expected, the fusion protein NLS–GFP, consisting of the basic candidate NLS sequence (amino acids 116–146) fused to GFP, was clearly restricted to the nucleus (Fig. 7A), indicating that this sequence was sufficient for nuclear translocation of Ca IRF-1. Two other constructs, ΔC–GFP and ΔN–GFP, also localized to the nucleus, whereas M–GFP and ΔN–GFP localized to the cytoplasm and nucleus, respectively. The localization patterns of the other constructs are summarized in Fig. 5. The subcellular distribution of Ca IRF-1 is shown in Fig. 6. (A) Nuclear localization of wild-type Ca IRF-1 tagged with GFP. CAB cells and CO cells were transiently transfected with plasmid GFP or IRF-1–GFP for 24 h, and then examined using a fluorescence microscope. PI staining shows the nuclei of cells transfected with IRF-1–GFP. (B) Observations over time of subcellular localization of Ca IRF-1 in CAB cells and CO cells when transiently transfected with IRF-1–GFP.
(GFP fused to the CalRF-1 N-terminus, including the DBD and potential NLS) and ΔN–GFP (GFP fused to the C-terminus of CalRF-1, including the potential NLS but lacking the DBD), were localized to the nuclear compartment (Fig. 7A). These data collectively support the conclusion that the amino acid sequence 116–146 is a functional NLS of CalRF-1.

Fig. 7 Identification of functional NLSs responsible for nuclear localization of CalRF-1. CAB cells and CO cells were transiently transfected with nine plasmids encoding CalRF-1 deletion mutants fused to GFP, or deletion mutants further modified using site-directed mutagenesis. Micrographs show GFP signal in cells transfected with NLS–GFP (116–146), ΔC–GFP (1–146), ΔN–GFP (116–289), ΔNLS–GFP (1–115 plus 147–289), N–GFP (1–115), C–GFP (147–289), M–GFP (1–72), NA1–GFP (K95, K97, K101 changed to A), or NA2–GFP (K95, K97, K101 changed to A and K75, K78, R82 changed to A).
The data in Fig. 7A show that amino acids 116–146 are sufficient to drive nuclear translocation of CaIRF-1. However, complete deletion of this portion in the ΔNLS-GFP fusion protein did not impair its nuclear localization (Fig. 7B). This finding indicated that, apart from the NLS within amino acids 116–146, there might be one or more additional NLSs in CaIRF-1. In order to test this hypothesis, two mutants, N–GFP (GFP fused to the DBD of CaIRF-1, spanning amino acids 1–115) and C–GFP (GFP fused to the C-terminus of CaIRF-1, encompassing amino acids 147–289), were constructed by deletion of the identified NLS from ΔC–GFP (1–146) and ΔN–GFP (116–289), respectively. When CAB and CO cells were transiently transfected with both CaIRF-1 mutants, C–GFP fusion protein, like wild-type GFP protein, showed diffused localization throughout the cell, while N–GFP was detected primarily in the nucleus (Fig. 7B). An additional mutant, M–GFP (GFP fused to the N-terminal 72 residues of CaIRF-1), also showed cytoplasmic localization (Fig. 7C). These data indicated that, besides the NLS located immediately C-terminal of the DBD (amino acids 116–146), another previously unidentified NLS exists within the DBD of CaIRF-1, more specifically within the region defined by amino acids 73–115, as this sequence must be responsible for nuclear retention of ΔNLS–GFP.

Upon further examination of the region between amino acids 73 and 115, the presence of five lysines (K75, K78, K95, K97, and K101) and three arginines (R82, R107, and R110) was noted. In order to determine the contribution of specific amino acids within this basic region to the nuclear accumulation of N–GFP, site-directed mutagenesis was performed. Two new constructs were obtained, plasmid NA1–GFP, in which three basic amino acids in N–GFP (K95, K97, and K101) were replaced with alanines, and NA2–GFP, in which six basic amino acids (K95, K97, K101, and K78, K78, R82) were changed to alanine. As shown in Fig. 7C, NA1–GFP was localized throughout the cell, albeit more fluorescence was apparent within the nucleus. NA2–GFP had a distribution pattern similar to that of wild-type GFP. These results indicate that the six basic amino acids within the region 73–115 can contribute to the nuclear localization of CaIRF-1.

4. Discussion

Significant progress has been made in the past several years in understanding the molecular mechanisms underlying fish innate antiviral responses [31]. CAB cells exposed to UV-inactivated GCHV have been used as a model for identification and isolation of virally induced genes, supported on the finding that UV-inactivated GCHV is more effective in inducing CAB IFN than active GCHV [39,40]. Subsequently, many genes involved in the innate IFN antiviral response have been cloned, including signal recognition factors TLR-3 and RIG-I, signal transduction and modulation factors such as STAT-1 and IRF-7, antiviral effectors such as Mx-1, Mx-2, Viperin, IFI56, PKR, and PKR-like, and other important IFN inducible genes such as IFI58, ISG15-1, ISG15-2, and USP18 [31 and unpublished data]. In the current study, a newly characterized IRF, most similar to zebrafish IRF-1 (76.2% identity in the whole sequence and 95.6% identity within the DBD) (Fig. 1), has been identified in CAB cells. The locus encoding this new IRF has organization and exon sizes characteristic of all members of the IRF-1 family (Fig. 3). Therefore, it is suggested that the IRF characterized herein should be considered a bona fide IRF-1, and termed CaIRF-1. CaIRF-1 can be induced by virus, CAB ICS, and Poly I:C, further supporting the notion that fish have a complete IFN system similar to that in mammals.

It is believed that IRF family factors have arisen by gene duplication and further diversification. The N-terminal 115 amino acids, representing the DBD, are the most conserved region in members of this family. The DBD contains five invariant tryptophan residues essential for its tertiary structure, which has been determined to be a helix–turn–helix motif based on nuclear magnetic resonance measurements [41]. The structural details that result in DNA binding by the DBD of IRF factors have been determined by X-ray crystallography [13]. To date, there are at least four members of the IRF family that have been identified in fish, namely IRF-1, IRF-2, IRF-3, and IRF-7 [24–29]. All the data collected to date suggest that each fish IRF family member is a true orthologue of its mammalian counterpart, although alternatively spliced variants found in some IRF family members can complicate the assignment of orthologous pairs. It is worthy of note that recent RFLP analysis suggests the existence of two turbot IRF-1 genes that might represent two different loci [29]. In the present study, comparison of genomic organization showed that all IRF-1 genes, including CaIRF-1 and pufferfish IRF-1, have identical disposition of exons and introns (Fig. 3) [25]. Although fish IRF-1 genes have a very compact size as a result of short introns, the sizes of exons are similar among all IRF-1 genes studied. This result further indicates that vertebrate IRF-1 genes are evolutionarily conserved.

CaIRF-1 also exhibited expression patterns that are similar to those of its mammalian counterparts, including ubiquitous and constitutive expression in uninfected CAB cells and fish organs, and upregulation in CAB cells stimulated by virus infection, CAB ICS, and Poly I:C treatment (Fig. 4). It is well known that UV-inactivated GCHV, but not active GCHV or Poly I:C, is able to induce high titers of IFN in CAB cells [39]. Based on this observation, the stronger induction of CaIRF-1 by UV-inactivated GCHV and CAB ICS than by FBV or Poly I:C may indicate that CaIRF-1 is predominantly responsive to IFN. Consistent with this idea, a recent sequence analysis of the 5′ flanking region of a rainbow trout IRF-1 gene has identified a structure that is typical of IFN-induced gene promoters, and this structure has been shown to be capable of driving the expression of a luciferase reporter upon induction by Poly I:C [42].

This study is not the first characterization of fish IRF-1, but it does represent the most comprehensive effort to date to describe its gene structure and the determinants for its nuclear localization. Like mammalian IRF-1, CaIRF-1 resides in the nuclear compartment (Fig. 6A, B), and its distribution is not altered by virus infection or Poly I:C treatment (data not shown). Because the size of the fusion protein IRF-1–GFP is well beyond the theoretical diffusion limit for the nuclear pore [43], nuclear import and retention of CaIRF-1 implies the existence of nuclear import signals. A series of localization experiments on CaIRF-1 mutants fused to GFP have provided direct evidence that the predicted bipartite NLS, located within the sequence YKKDKRKPKGRDSRRRVKALSHSVKK [44], is sufficient for nuclear translocation of
CoIRF-1 (Fig. 7). Apart from the parallels that the discovery of this NLS draws between fish and human IRF-1 proteins, the present study also revealed significant differences in the mechanisms that fish and mammalian IRF-1 utilize to enforce nuclear localization. As reported previously, human IRF-1 relies exclusively on the classic bipartite NLS for nuclear localization, which corresponds in location and approximate amino acid composition to the NLS found here in the region between amino acids 116 and 146 of CoIRF-1. In contrast, nuclear retention of the mutant ANLS–GFP of CoIRF-1 indicates the presence of additional nuclear localization determinants in the fish IRF-1. Further analyses of mutants of CoIRF-1 revealed a new NLS, located in a region within the DBD (positions 73–115), and including two amino acid sequences rich in basic residues ("KKDKSKINK101" and "KKWKANFR115"). Some of these basic residues were found, by mutation analyses, to be important for full activity of the novel NLS. Therefore, the current study has revealed a new nuclear localization motif in the DBD of fish IRF-1, which thus far appears to be a situation unique among vertebrate IRF-1 family members. Other unique properties have been found in the IFN system of fish. For example, apart from a PKR gene, fish has a PKR-like gene, which is not structurally identical to mammalian PKR [44]. Crucian carp IFIS6, contrary to human IFIIS6, is not the most highly responsive gene during the IFN response, and displays unique induction mechanisms in CAB cells treated with UV-inactivated GCHV [31,45].

Considerable diversity is observed in sequence and localization of the NLS among mammalian IRFs [19]. For example, IRF-4 and IRF-9 have a classic bipartite NLS that resides in the N-terminal DBD [20], whereas IRF-5 has two NLSs, one residing in the DBD and a second one in its C-terminus [19]. More interestingly, IRF-3 has an NLS that is composed of just two amino acids, 77KR, and is also located within the DBD [21]. In the current study, a second NLS was also found within the DBD of CoIRF-1. Five out of six basic residues included in this second NLS (K75, K78, R82, K95, K101, but not K97) are conserved in mammalian IRF-1 proteins. In mammalian IRF-1, however, this basic region does not function as an NLS. This discrepancy implies that K97 of CoIRF-1 might be essential for the function of its second NLS, wherein the six basic amino acids might cooperate to drive CoIRF-1 to the nucleus.

Acknowledgments

This work was supported by grants from the National Basic Research Program of China (2004CB117403), the National Natural Science Foundation of China (30471333, 30671617), and the Innovation group project of Hubei Province (2004ABC005).

References


